Segregating inclusive, prompt and non-prompt production of J/ψ at the LHC energies using machine learning

Suraj Prasad
Indian Institute of Technology Indore, India
Email: Suraj.Prasad@cern.ch

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

Outline

- Introduction
- Quarkonia
- Topological production of J/ψ
- Inputs to the machine
- Model parameters
- Model performance
- Results
- Summary

Outline

- Introduction
- Quarkonia

Big Questions:

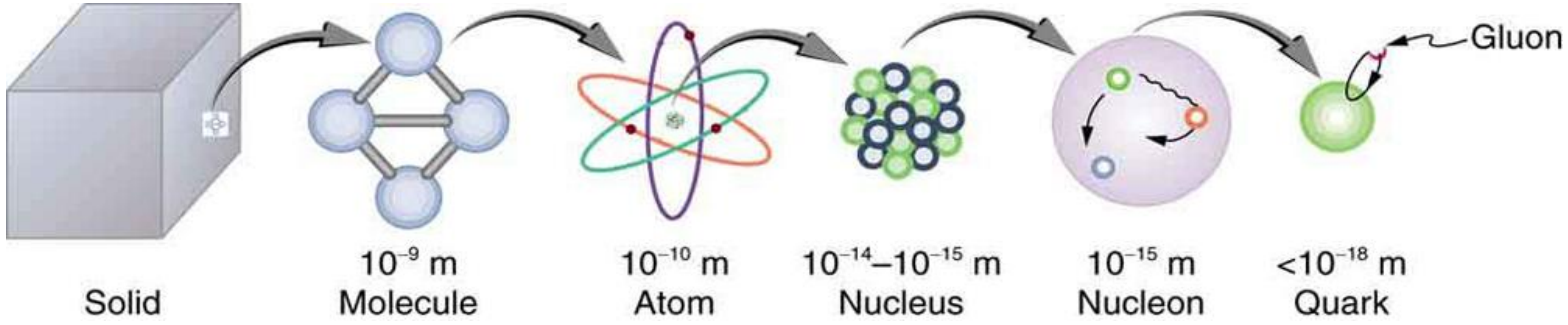
What is the universe made of?

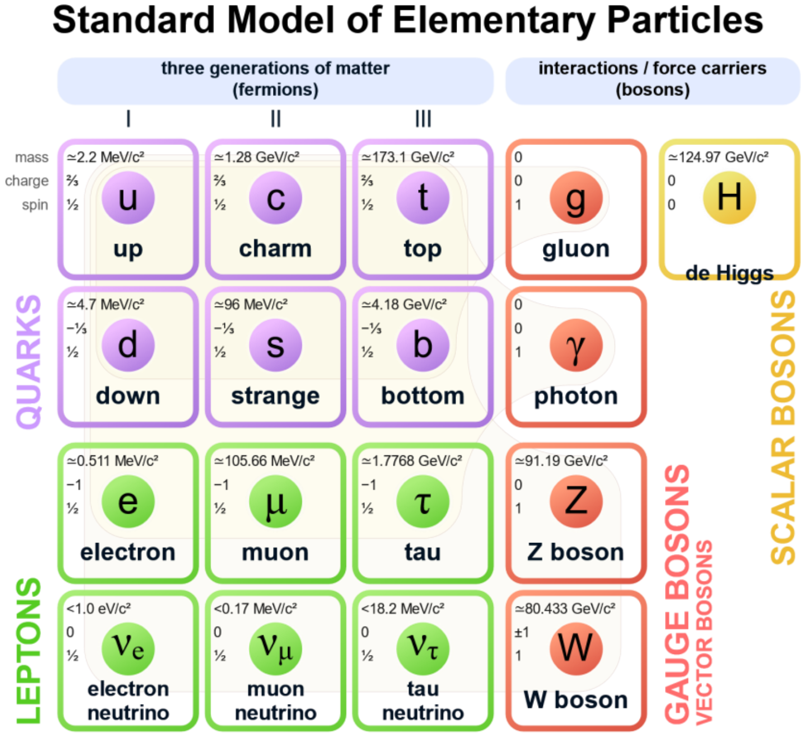
How does it work?

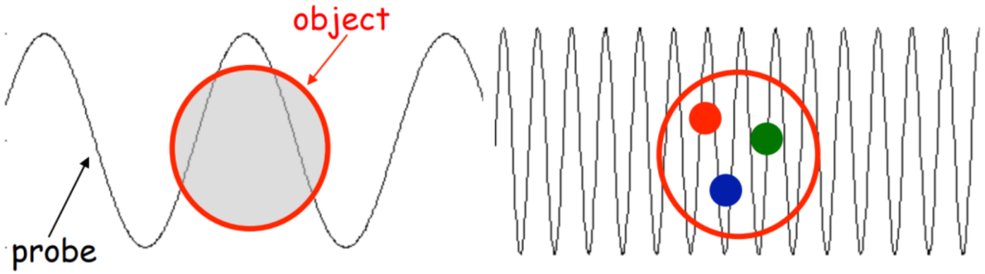
How did it evolve?

- Model performance
- Results
- Summary

Constituents of Matter



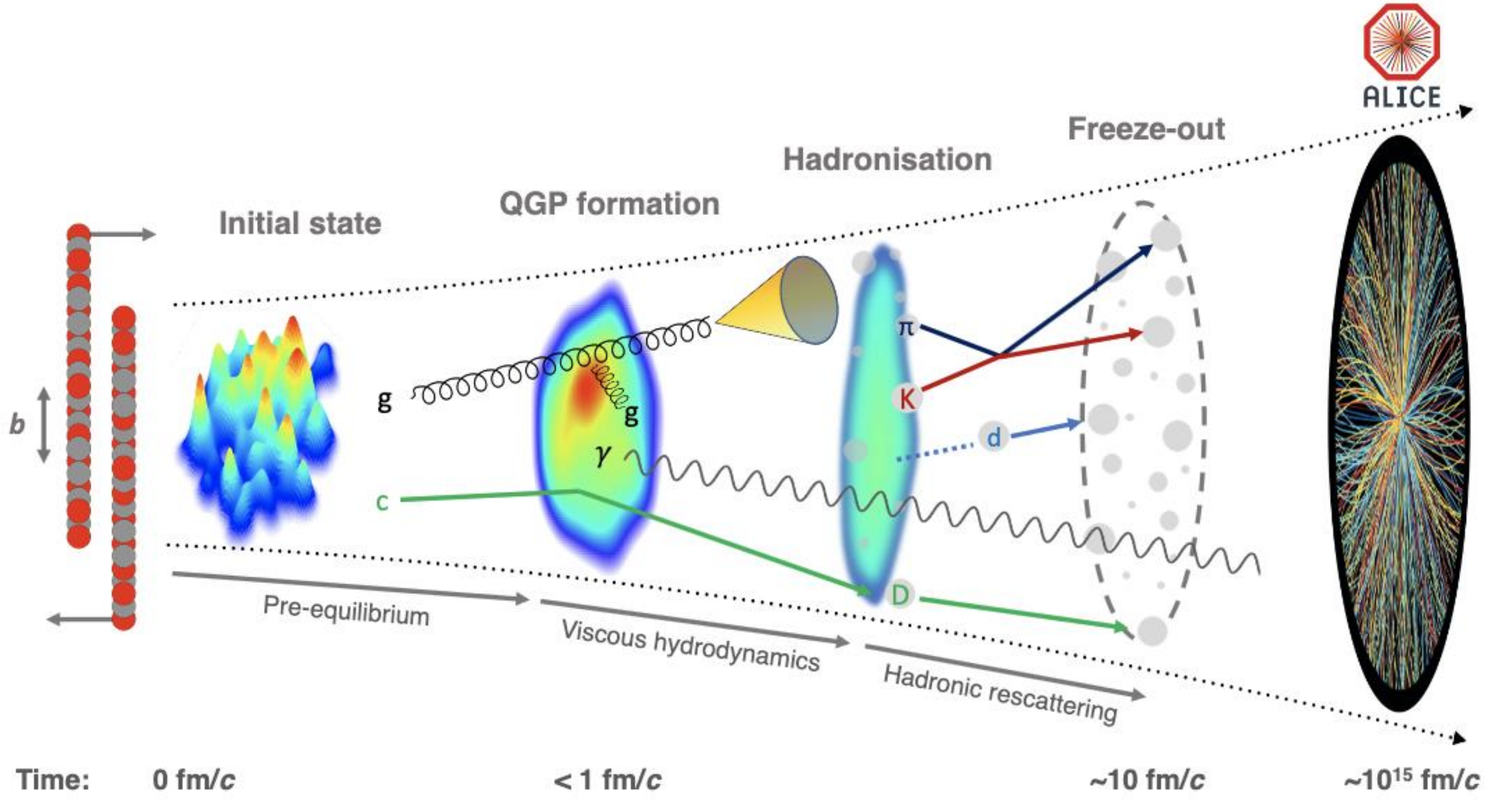


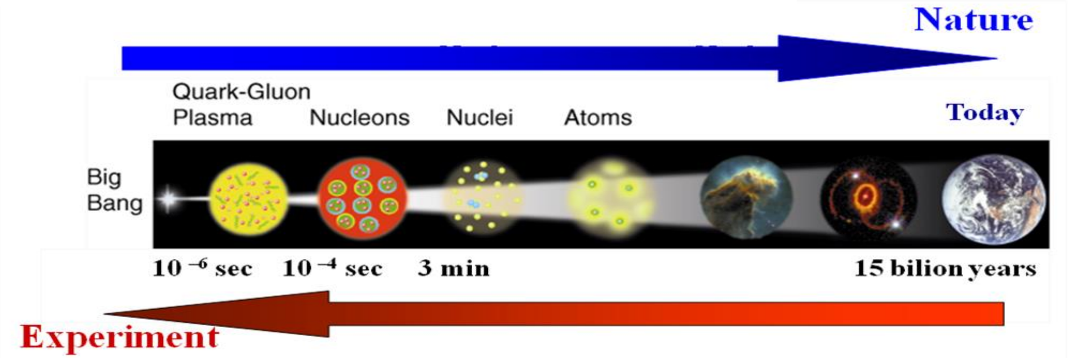


- Energy is related to wavelength by de Broglie's formula: $p = h/\lambda$
- To probe inside smaller objects, we need higher energy

[1] R. Sahoo, "Relativistic Kinematics", [arXiv:1604.02651 [nucl-ex]]

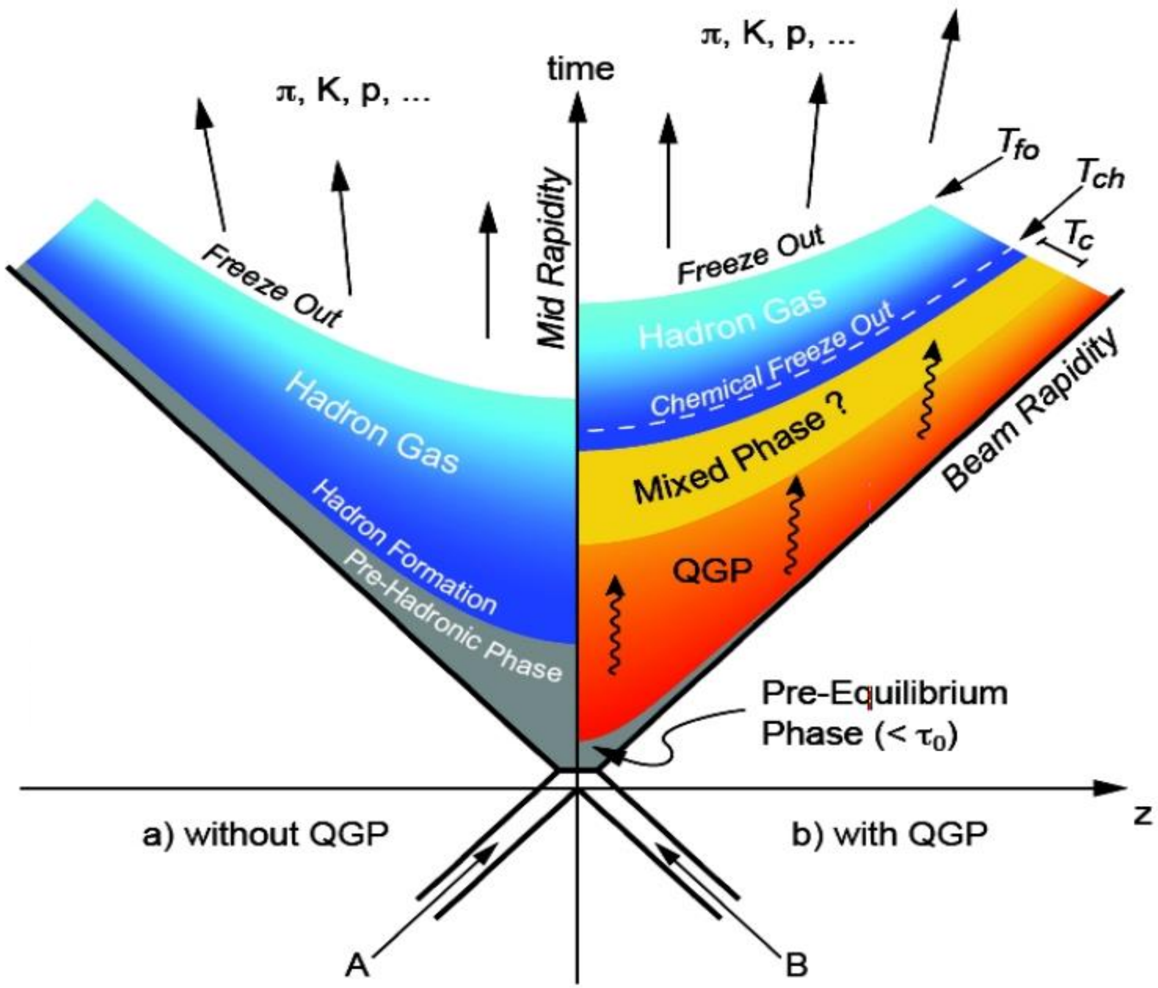
Space-time evolution in Collider Experiments





[1] R. Sahoo, AAPPS Bull. 29, 16 (2019).
[2] U. Heinz, Int. J. Mod. Phys. A 30, 1530011 (2015).

Space-time evolution in Collider Experiments



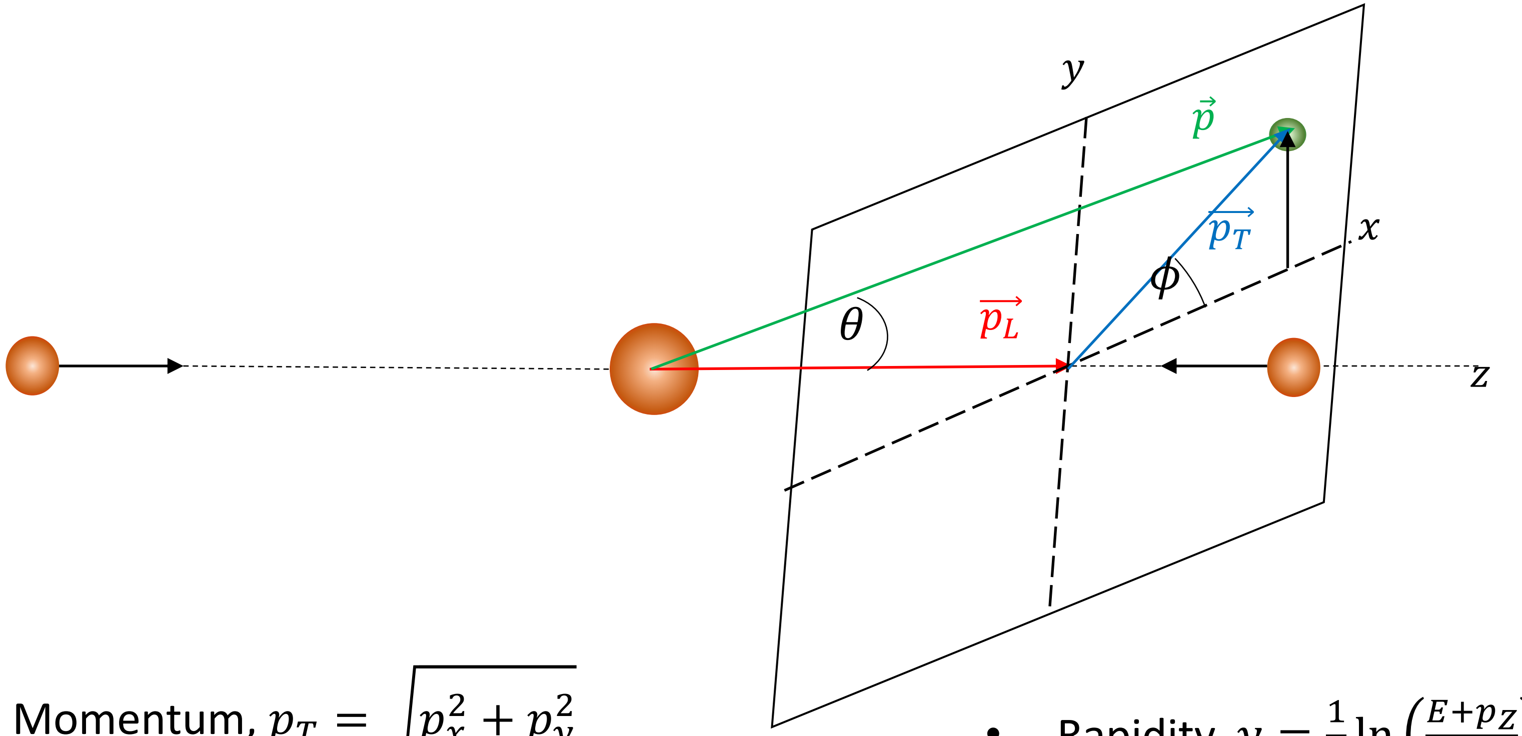
https://particlesandfriends.wordpress.com/2016/10/14/evolution-of-collisions-and-qgp/

- There are two possible scenarios of the space-time evolution in collider experiments depending upon the system size and the collision energy

- One assumes the formation of a deconfined thermalised state of the deconfined quarks and gluons known as the quark-gluon plasma (QGP) led by the pre-equilibrium phase and followed by a mixed phase and hadron gas phase (larger system size and denser partonic medium)

- Another scenario involves the prehadronic phase followed by the hadron gas phase (small collision scenario)

Coordinate System

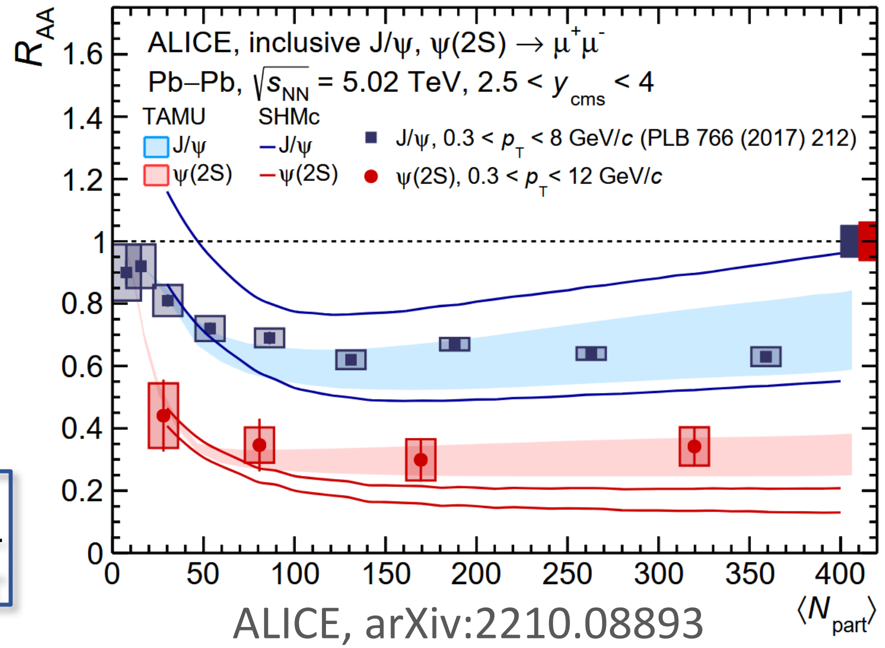


- Transverse Momentum, $p_T = \sqrt{p_x^2 + p_y^2}$
- Azimuthal Angle, $\phi = \tan^{-1}\left(\frac{p_y}{p_x}\right)$
- Polar angle, $\theta = \tan^{-1}\left(\frac{p_T}{p_z}\right)$
- Rapidity, $y = \frac{1}{2}\ln\left(\frac{E+p_Z}{E-p_Z}\right)$
- Pseudo-rapidity, $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$
- Every produced particle is represented in terms of their $(p_{\mathrm{T}}, \eta, \phi)$

[1] R. Sahoo, "Relativistic Kinematics", [arXiv:1604.02651 [nucl-ex]]

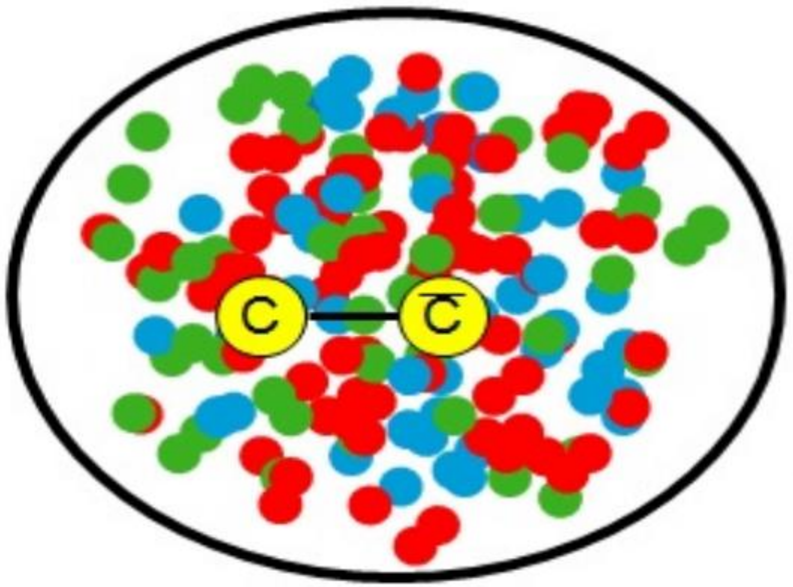
Quarkonia

- Quarkonia is a bound state of heavy quark and antiquark pairs
- Due to its heavy mass, quarkonia studies in QGP are important as it experiences the whole medium evolution
- Serve as the testing ground for QCD



ALICE, arXiv:2210.08893

$$R_{AA} = \frac{d^2N_{AA}/d\eta dp_T}{\langle N_{coll}\rangle\, d^2N_{pp}/d\eta dp_T}$$

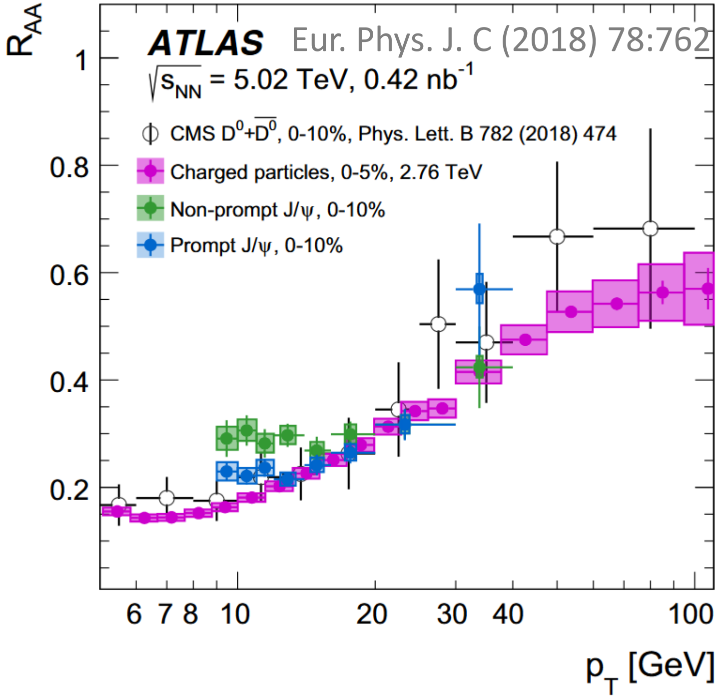


- Charmonia ($c\bar{c}$) and bottomonia ($b\bar{b}$)
- Suppression increased towards higher $\langle N_{part} \rangle$ due to the denser partonic medium: more screening

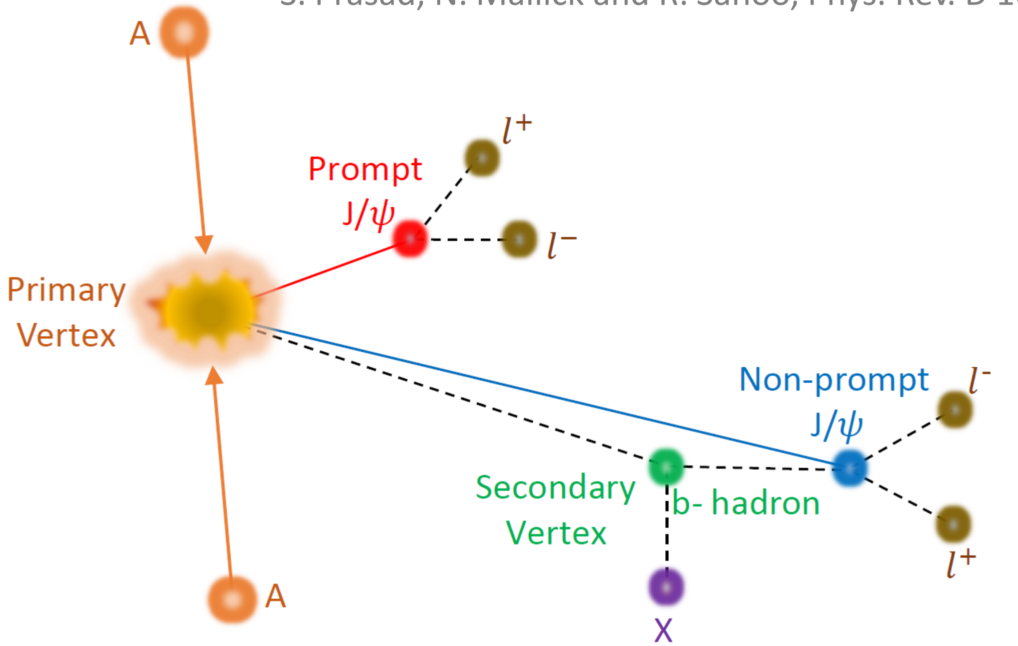
https://twiki.cern.ch/twiki/bin/view/ReteQuarkonii/ReteQuarkonii

Topological production of J/ψ

- J/ψ meson: Vector charmonium with lightest mass (3.096 GeV/c^2)
- In experiments, dileptonic channels are used to reconstruct J/ψ. (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 (Opportunity to study b-hadron)
- Prompt and non-prompt J/ψ are topologically different thus they both show different values of suppression

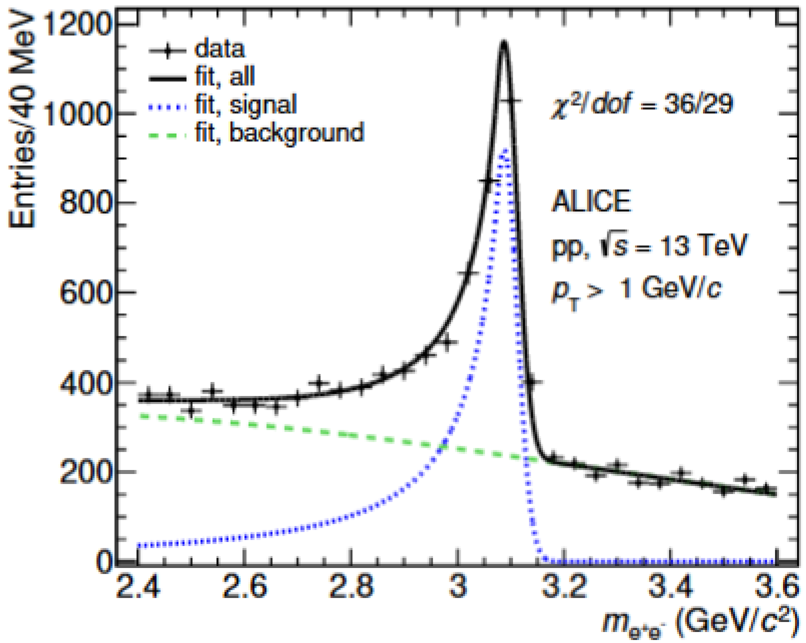


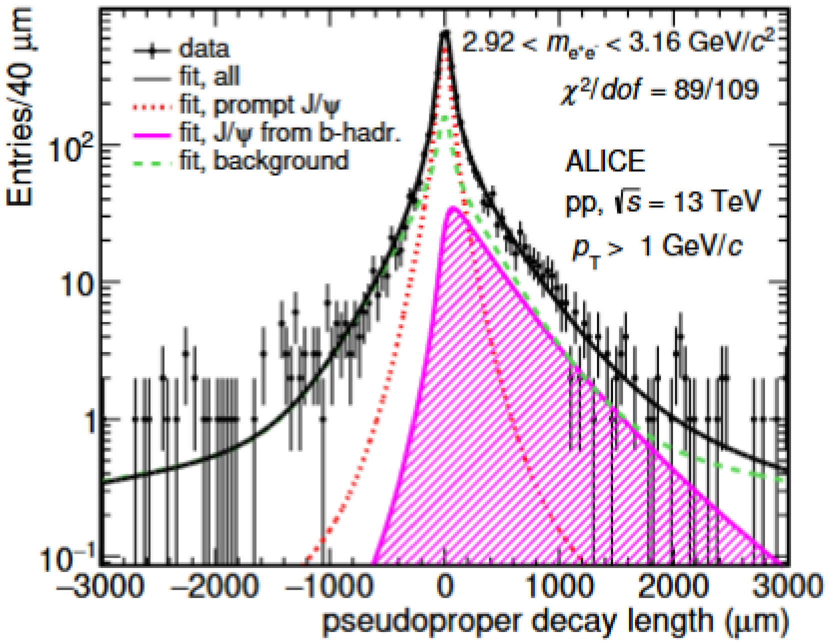
S. Prasad, N. Mallick and R. Sahoo, Phys. Rev. D 109, 014005 (2024)

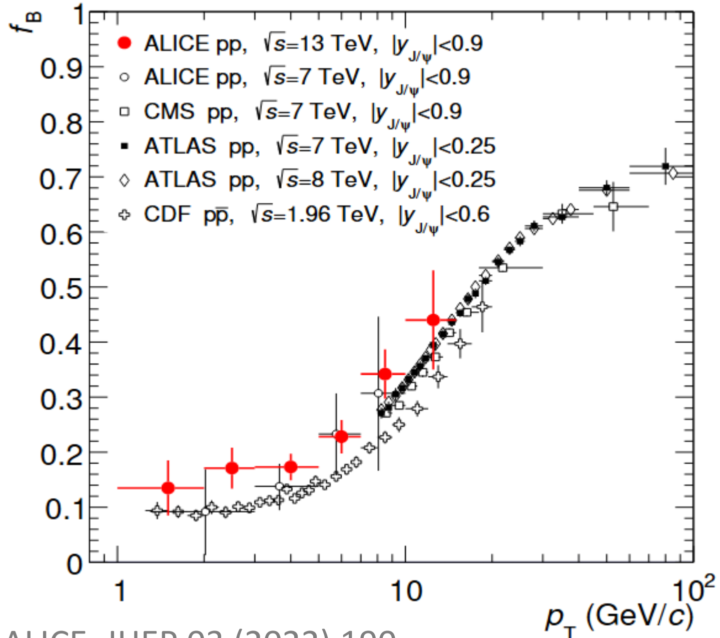


Experimental Procedure

- In experiments, the invariant mass of dilepton pairs are estimated: $M_{ee} = \sqrt{(E_1 + E_2)^2 - (|\vec{p_1} + \vec{p_2}|)^2}$
- Using the vertexing information from the detectors, the pseudoproper decay length ($c\tau$) is estimated: $c\tau = \dfrac{c\, m_{\mathrm{J}/\psi}\, \vec{L}\, .\, \vec{p_{\mathrm{T}}}}{|\vec{p_{\mathrm{T}}}|^2}$
- One performs a simultaneous fit to the invariant mass signal and pseudoproper decay length to obtain fraction of nonprompt yield (f_{B})
- For the fitting, the PDFs for prompt and non-prompt are usually taken from MC simulations





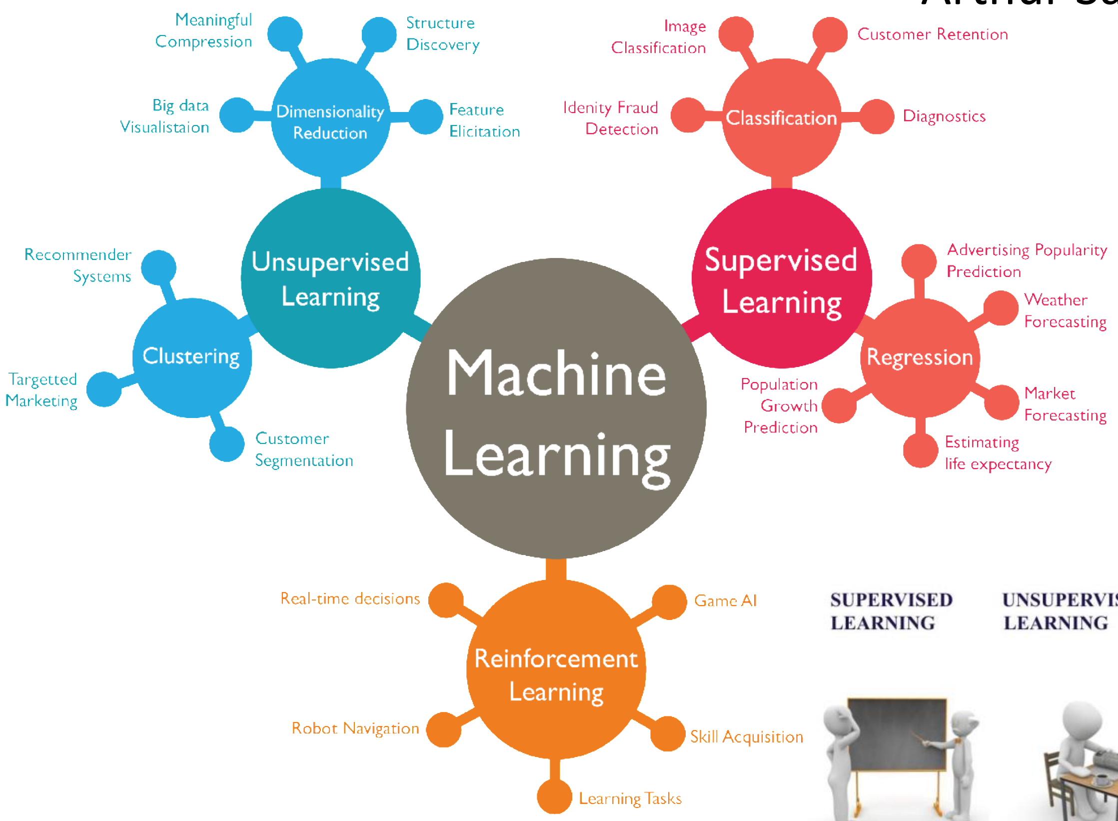


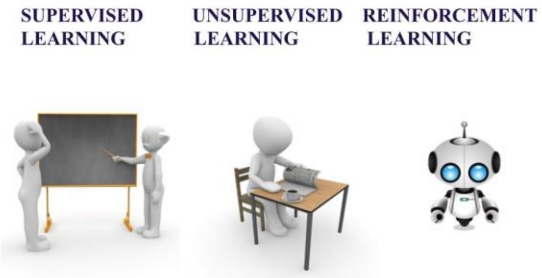
ALICE, JHEP 03 (2022) 190

Machine Learning

"Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed."

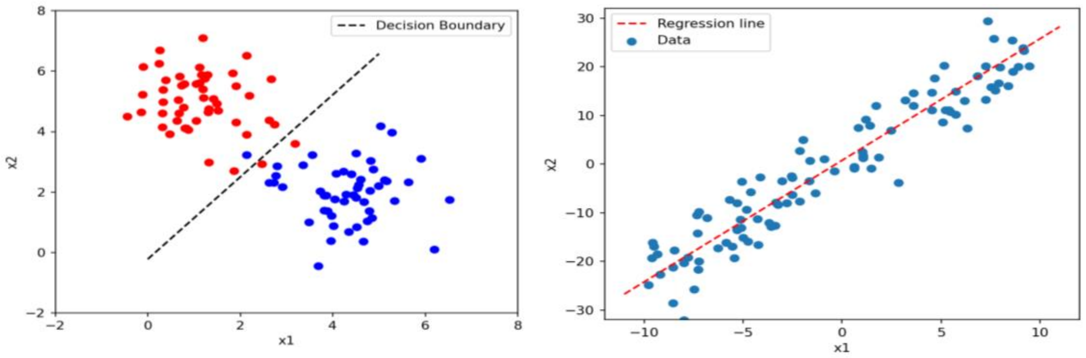
-Arthur Samuel, 1959

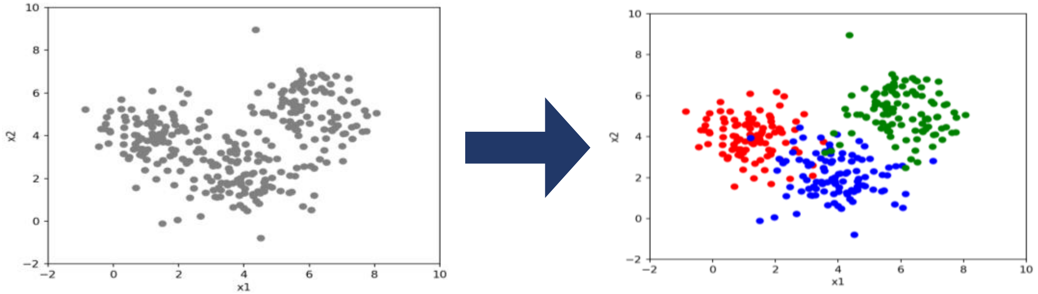




What it needs?

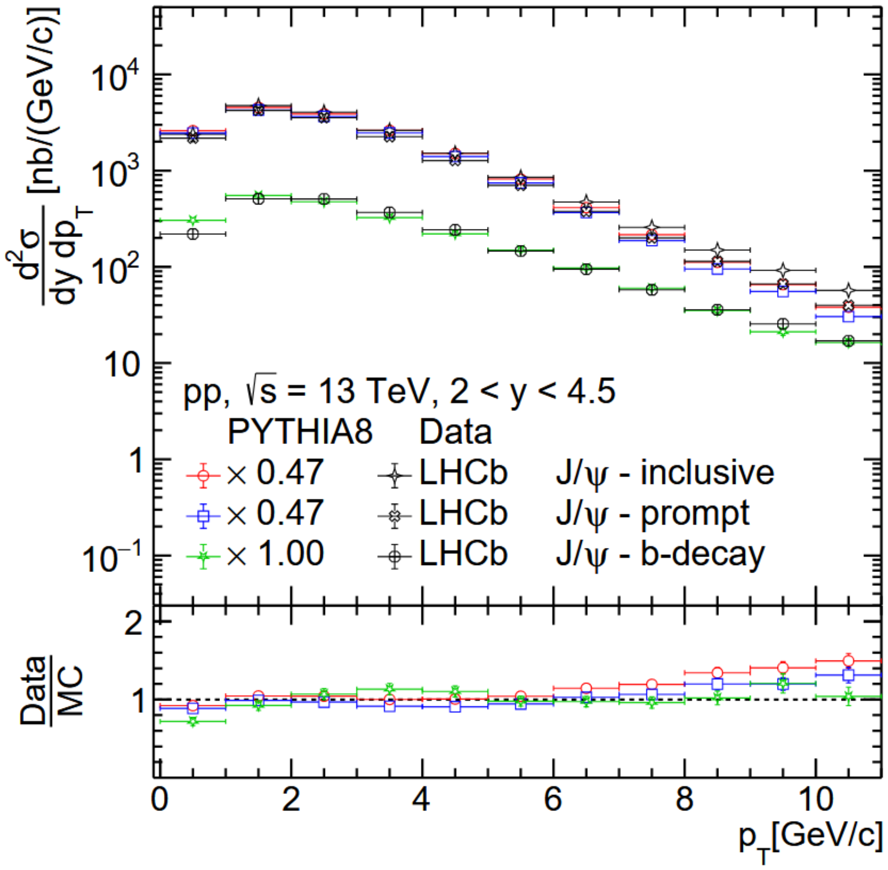
- Big data
- Smart algorithm (BDT, DNN, GNN etc.)
- Knowledge from data
- Tune the parameters (Optimise the model)
- Predict!!





Preparing the inputs

- PYTHIA8 is used as the MC generator to generate 20 billion minimum bias events for pp collisions at $\sqrt{s} = 13$ TeV using 4C-tune
- The coordinates of the primary vertex are randomised 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



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

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

$\vec{V}$ = Primary Vertex

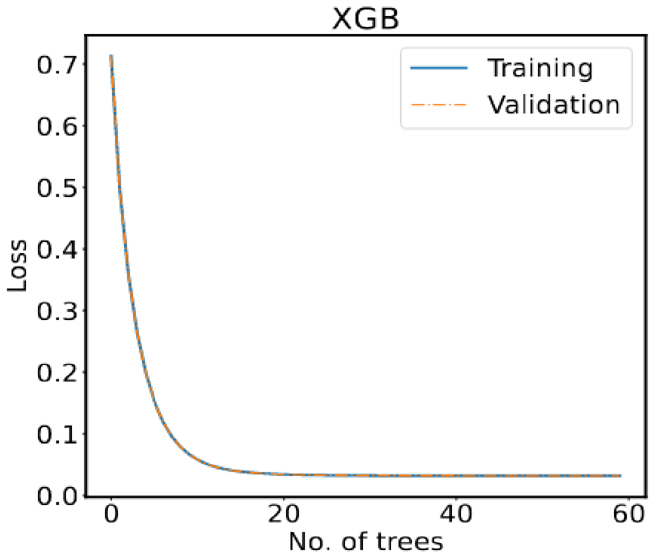
$\vec{S}$ = Reconstructed J/ψ decay vertex

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

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

Model parameters

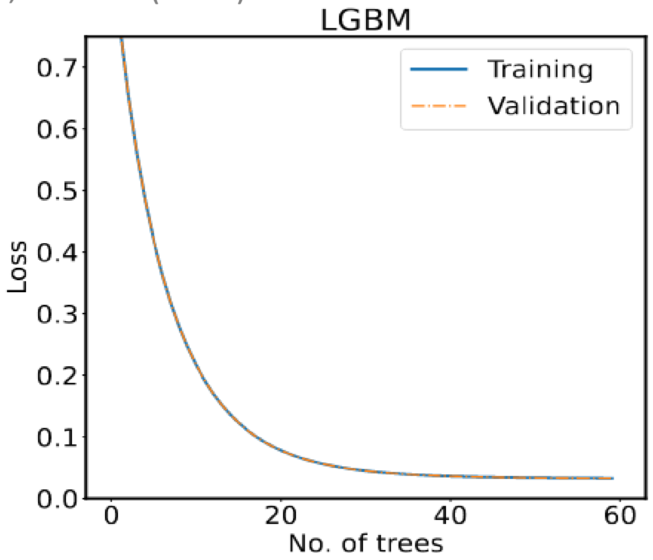
- Background : Prompt : Non-prompt = 20 : 10 : 1
- Classification models required to be trained on similar number of training instances → oversampling of data is done
- Dataset for Training : Testing : Validation = 81 : 10 : 9
- Parameters are chosen through a grid search method (Making an array of all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

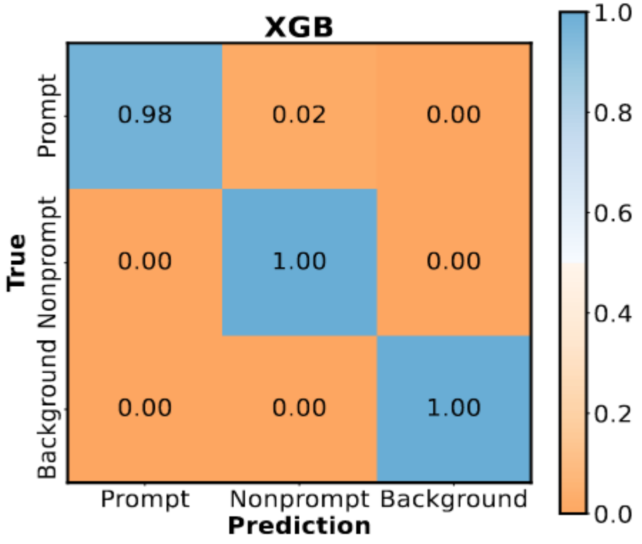


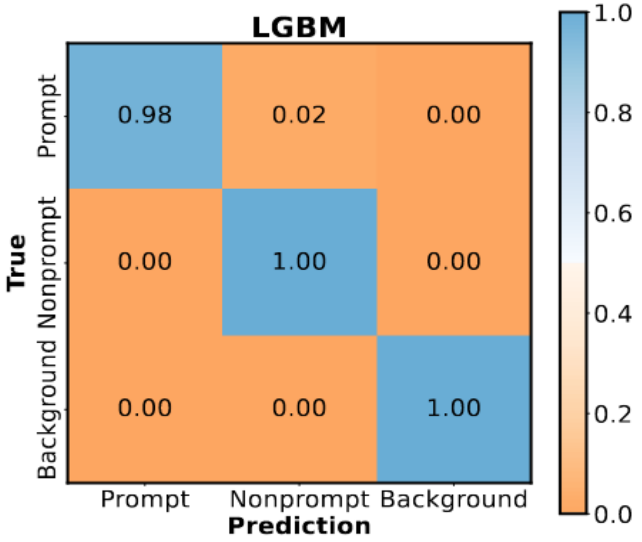
S. Prasad, N. Mallick and R. Sahoo, Phys. Rev. D 109, 014005 (2024)

	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*

- 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

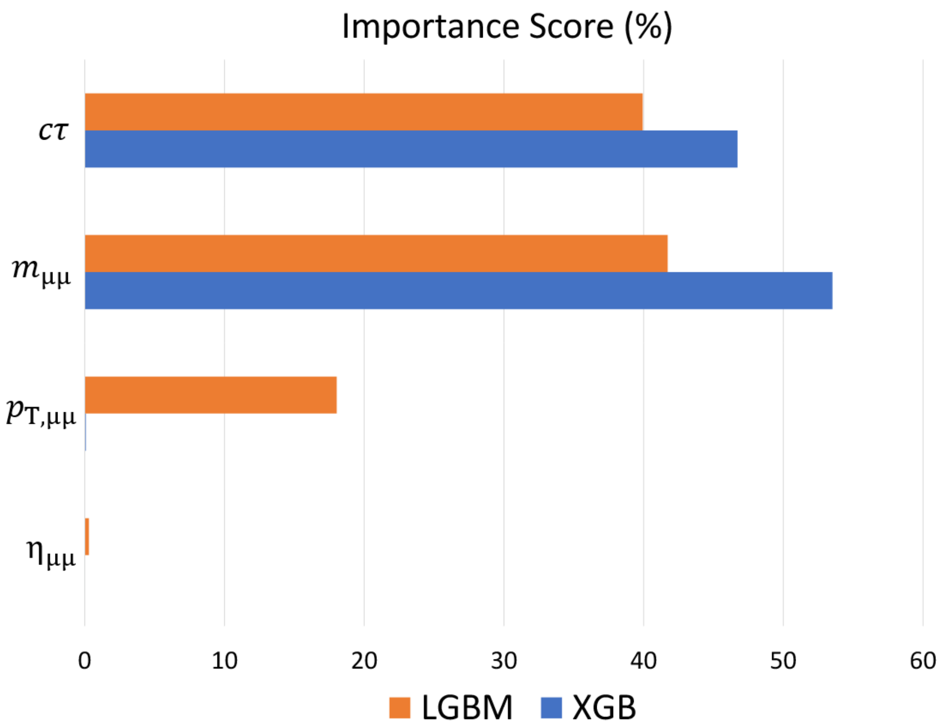






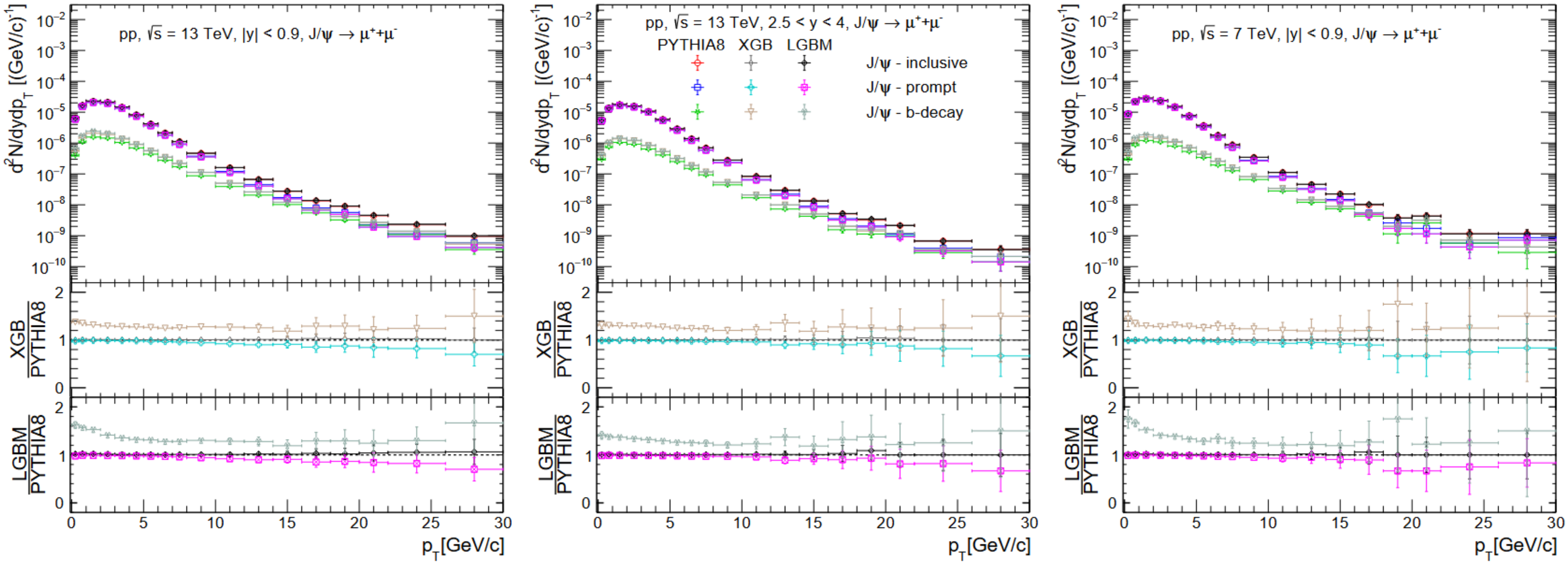
- Confusion Matrix talks about the mispredictions given by the model for each class
- Both XGB and LGBM perfectly separates the inclusive J/ψ from the uncorrelated background pairs
- Both models mispredict 2% of prompt J/ψ as the non-prompt → Raises non-prompt yield

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

Results: Transverse momentum spectra

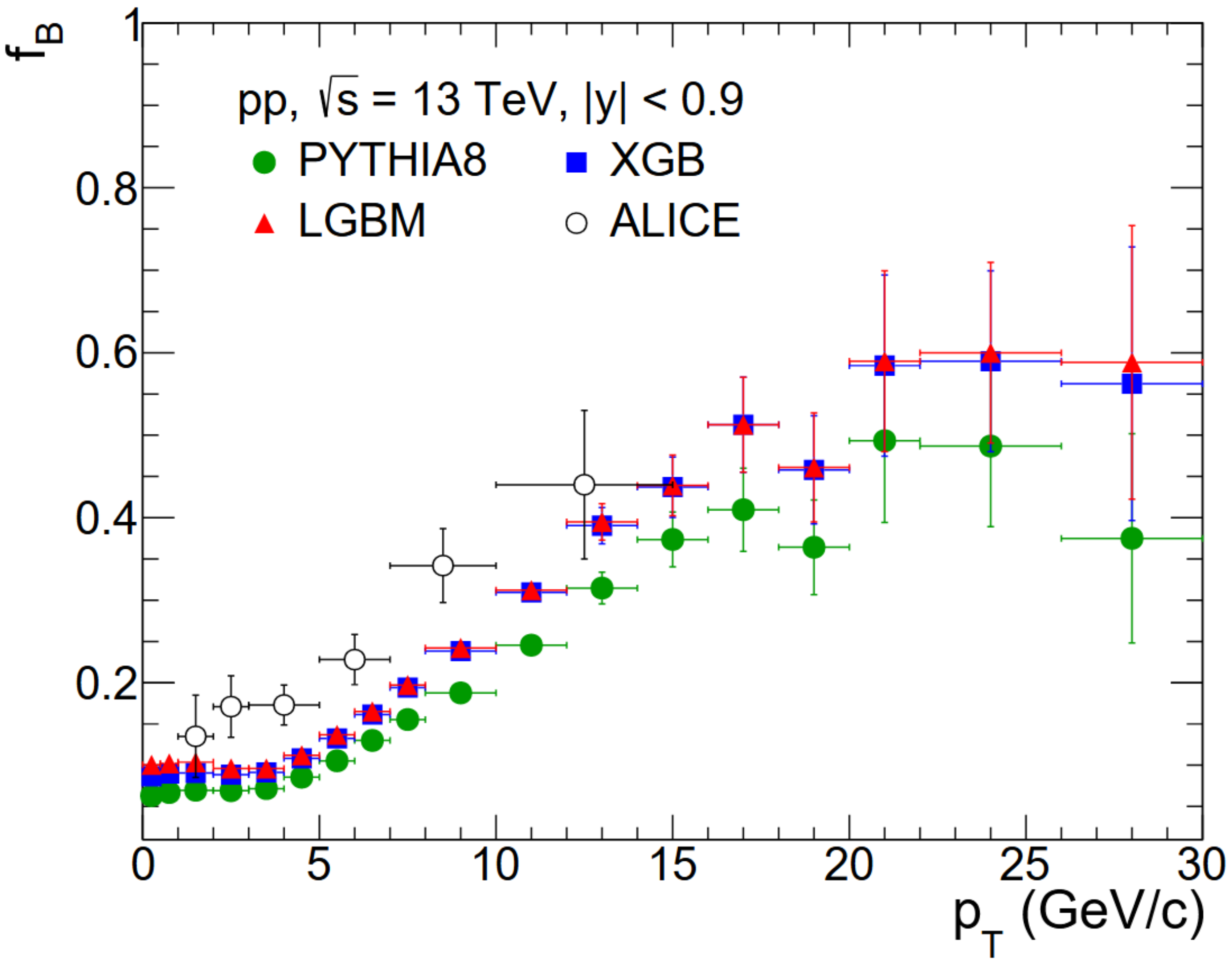


- Both XGB and LGBM give accurate predictions for p_{T}-spectra for inclusive and prompt-J/ψ both in mid and forward rapidity in pp collisions at $\sqrt{s} = 13$ TeV and 7 TeV

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

- 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

Results: Fraction of non-prompt J/ψ yield



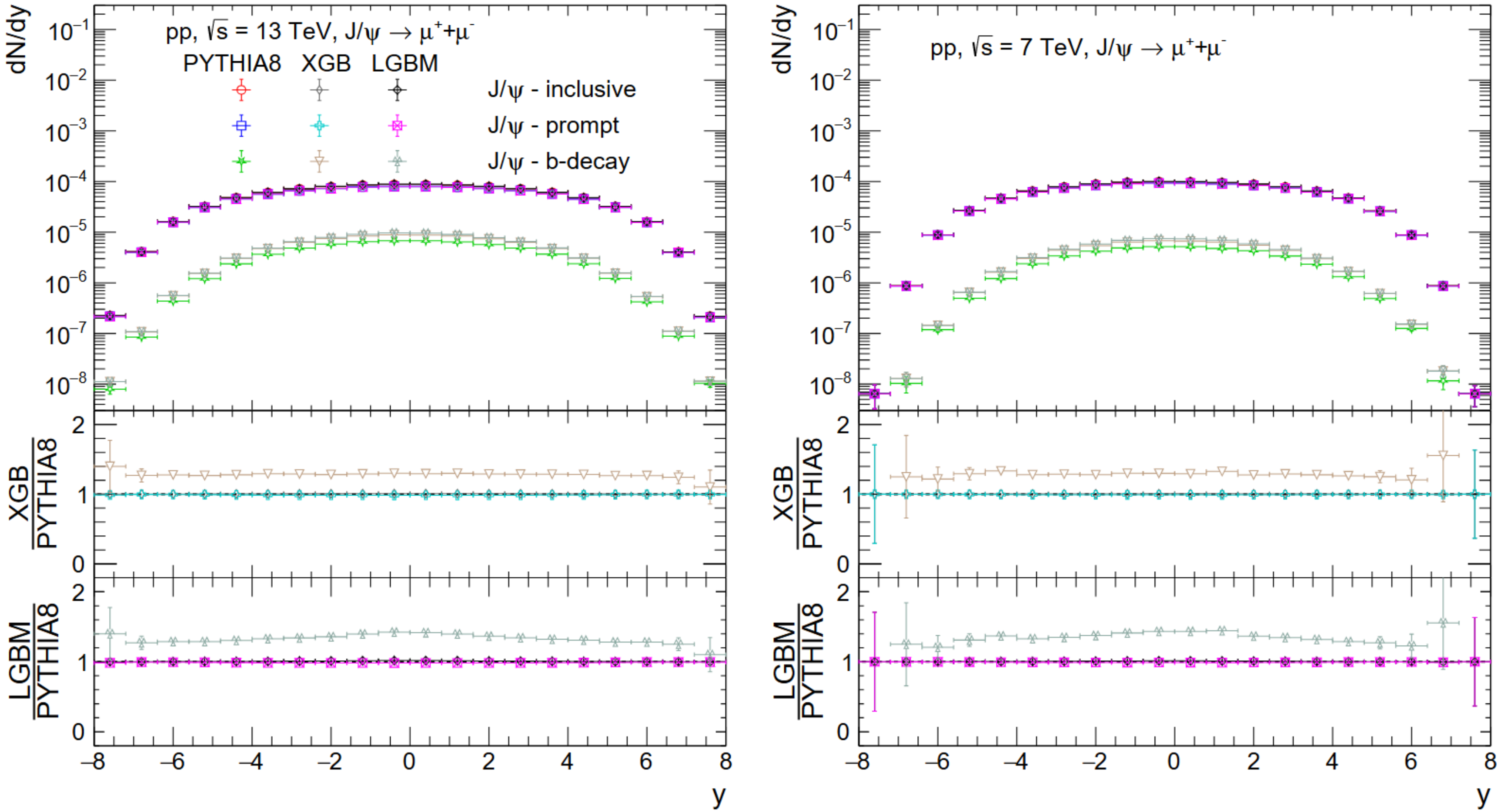
- f_{B} is the fraction of the non-prompt production (B-hadron decays)
- f_{B} increases with increase in p_{T} → The b-hadron production is favoured towards higher p_{T} compared to low p_{T}
- PYTHIA8 underestimates the experimental data following the similar trend
- Both XGB and LGBM overestimate PYTHIA8
- As this method does not require fitting, thus 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

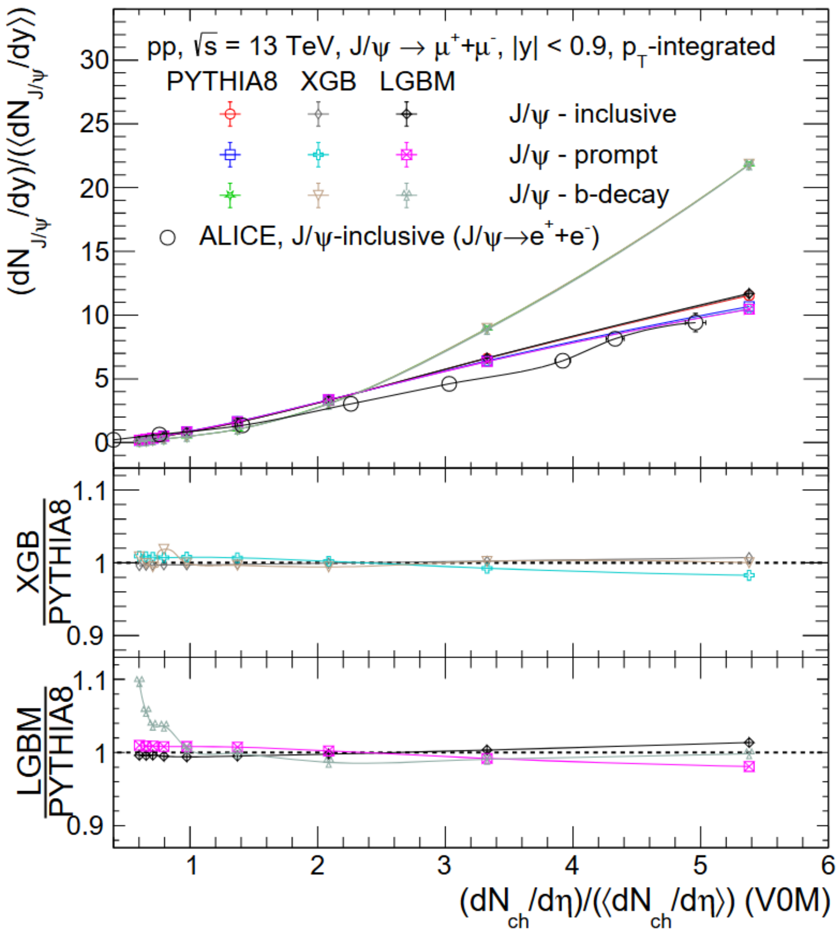
- Both XGB and LGBM give accurate predictions for rapidity spectra for inclusive and prompt-J/ψ in pp collisions at $\sqrt{s} = 13$ TeV and 7 TeV

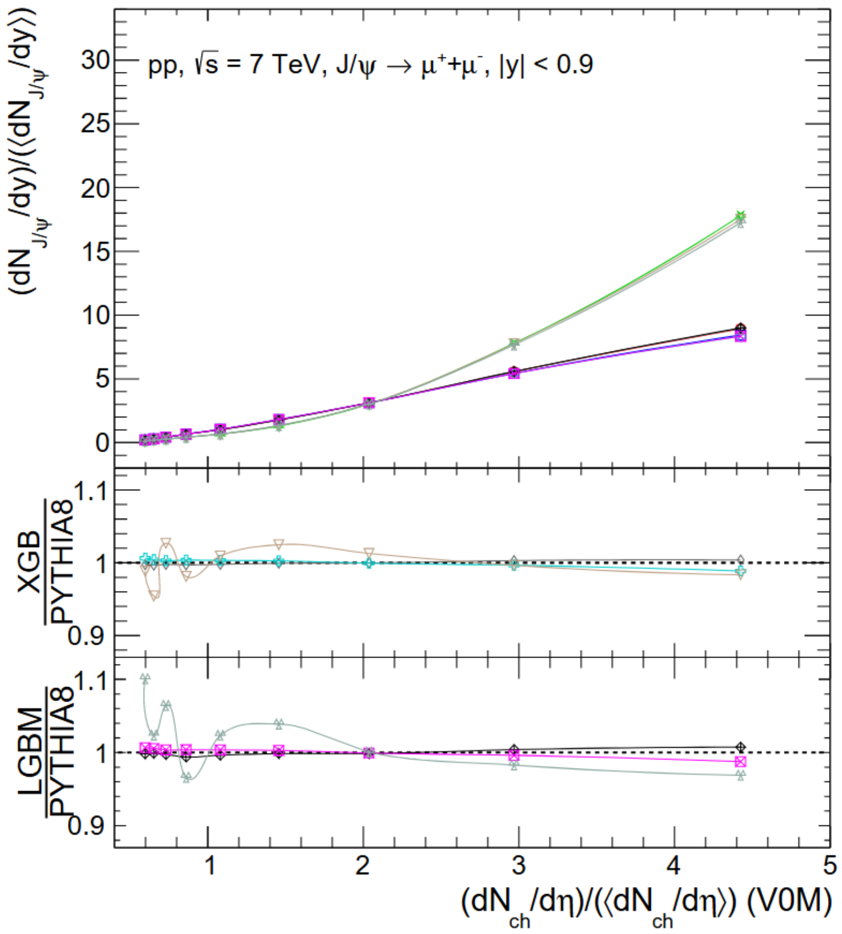
- The ML models overpredict the non-prompt J/ψ throughout rapidity region for both the collision energies



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 normalised 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 LGBM matches PYTHIA8 findings within 10% uncertainty

Summary

- We have used BDT based ML models such as XGBoost and LGBM to segregate the prompt, non-prompt and inclusive J/ψ production in pp collisions at $\sqrt{s} = 13$ TeV
- The models use the parameters, such as, pseudo-proper decay length ($c\tau$), invariant mass ($m_{\mu\mu}$), transverse momentum ($p_{T,\mu\mu}$), pseudorapidity ($\eta_{\mu\mu}$) of the dimuons as the input, which are accessible in the experiments
- The model almost achieves 99% overall accuracy
- The estimations for the prompt and inclusive J/ψ from the ML models match with the PYTHIA8 for the inclusive and prompt J/ψ
- Using this models, track label identification is possible, and it avoids the necessity of fitting of spectra
- The model is expected to work throughout the energy regime from RHIC to LHC and in heavy-ion collisions

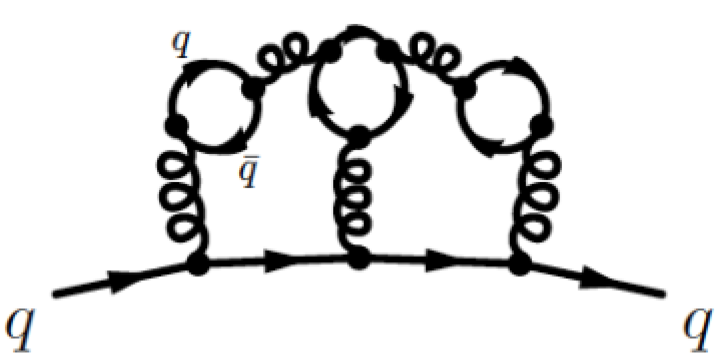
S. Prasad, N. Mallick and R. Sahoo, Phys. Rev. D 109, 014005 (2024)

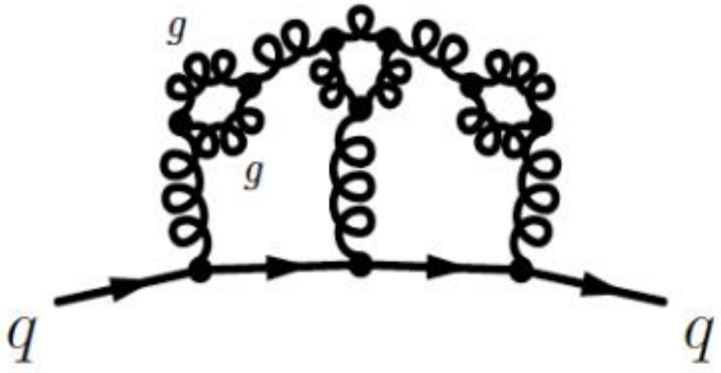
Thank you for your attention

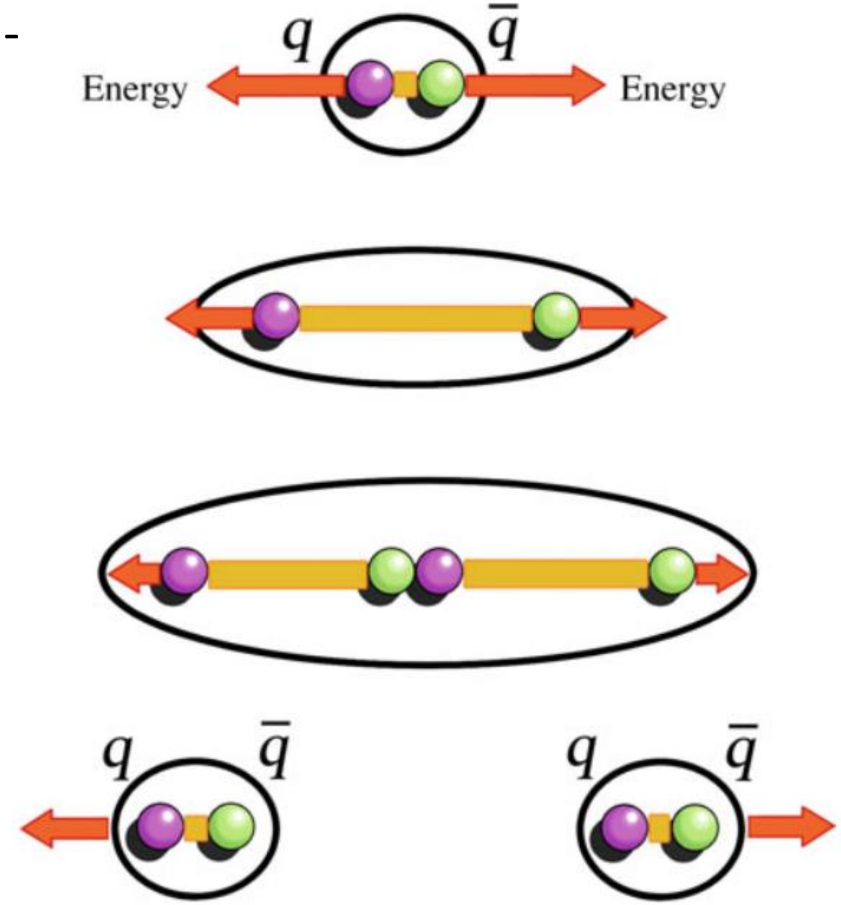
Backup

Strong Interaction

- Unlike QED, in QCD gluons have a color charge, which permits gluon-gluon interaction
- Color charges can't freely exist: Color confinement
- At high energies, α_s becomes smaller: Asymptotic freedom







Obertelli, A., Sagawa, H. (2021). Nuclear Physics and Standard Model of Elementary Particles. In: Modern Nuclear Physics. UNITEXT for Physics. Springer, Singapore

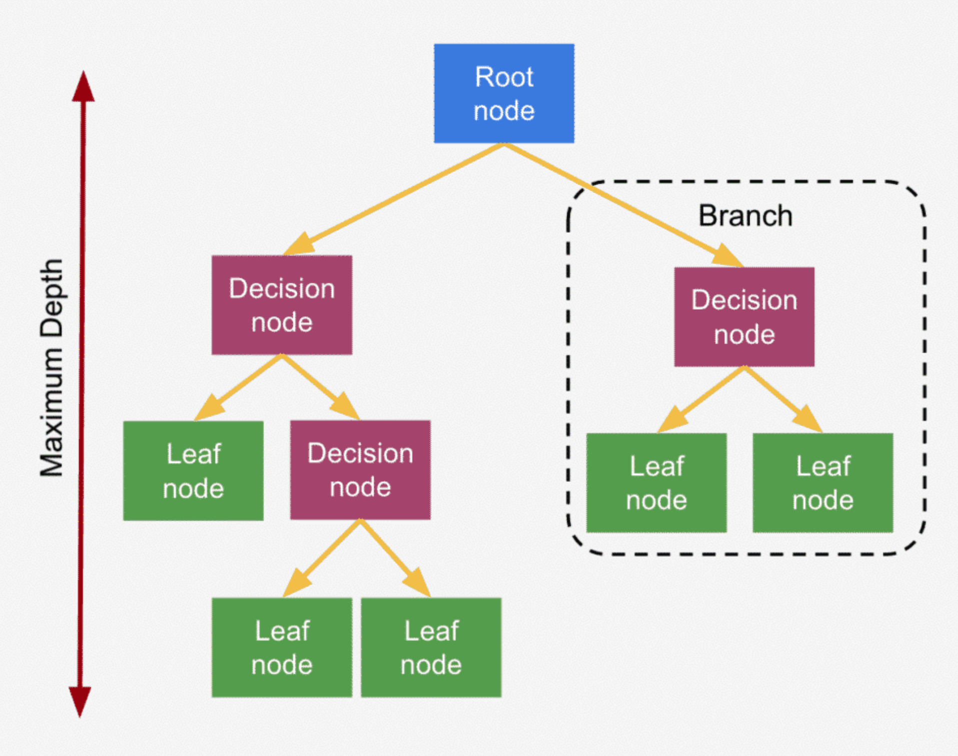
Gradient Boosting Machine

- 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)

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



- **Criteria of splitting:**

  Classification: Minimise the node impurity

  Regression: Minimise the MSE, MAE

MSE: Mean Squared Error
MAE: Mean Absolute Error

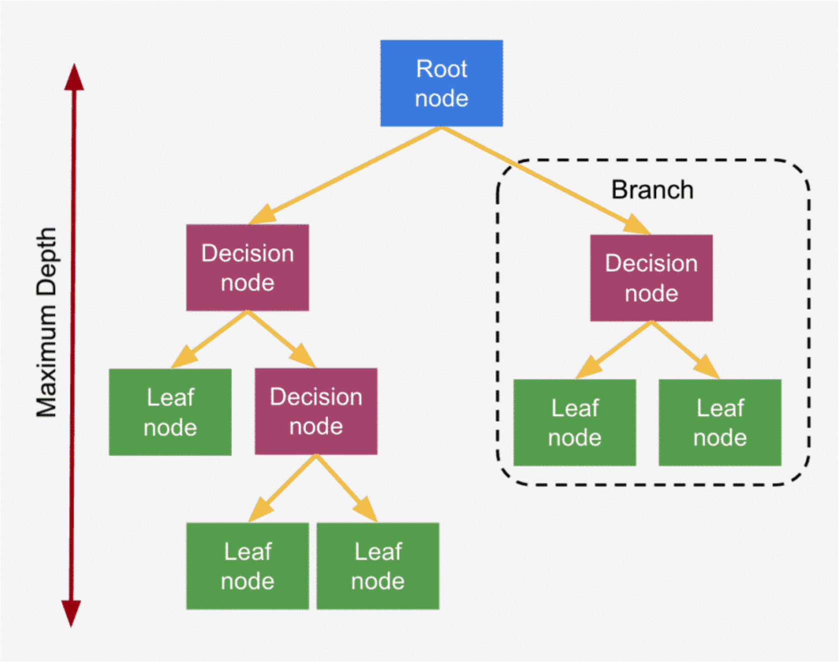
- Splitting continues till a preset (max_depth)
- **Boosting:** Building an additive forward staged model by combining the outcomes of all previous ones
- Boosting compensates the shortcomings
- Shortcomings are identified as the gradients
- Light Gradient Boosting Machine (LGBM): Leafwise splitting of tree, low memory use and supports parallel boosting
- Extreme Gradient Boosting (XGB): Advance version of Gradient Boosting that supports parallel tree boosting → Faster

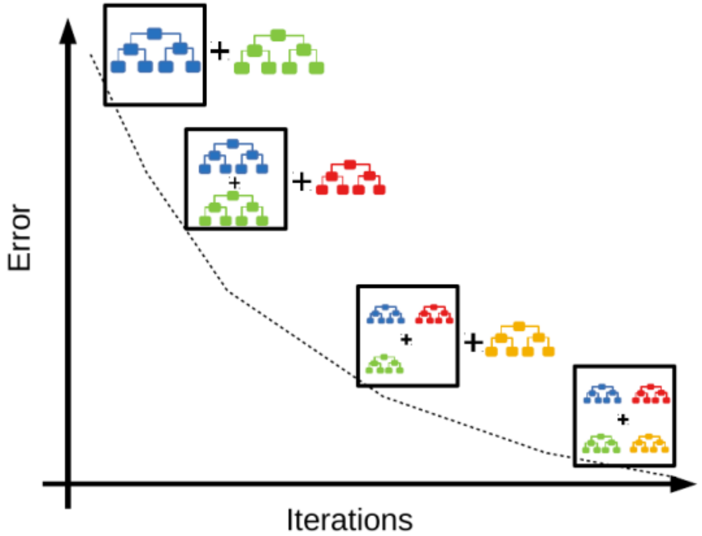
Gradient Boosting Machine

- **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

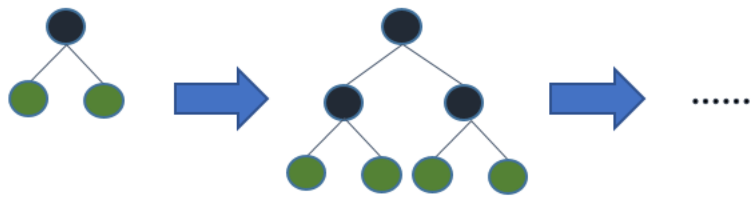




- Two Methods for making an ensemble of decision trees: Boosting and bagging
- **Bagging** method builds models in parallel using a random subset of data (sampling with replacement) and aggregates predictions of all models
- **Boosting** method builds models in sequence using the whole data, with each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

XGBoost

- Extreme Gradient Boosting

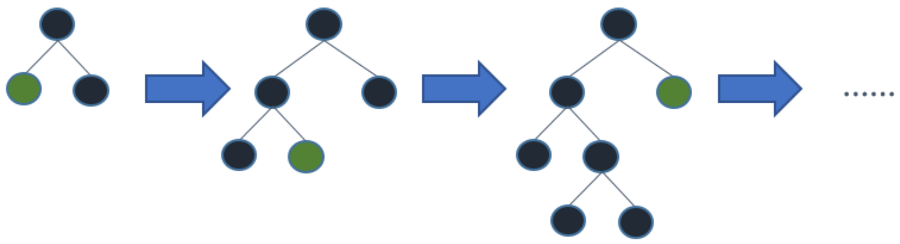


Level-wise tree growth

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

LightGBM

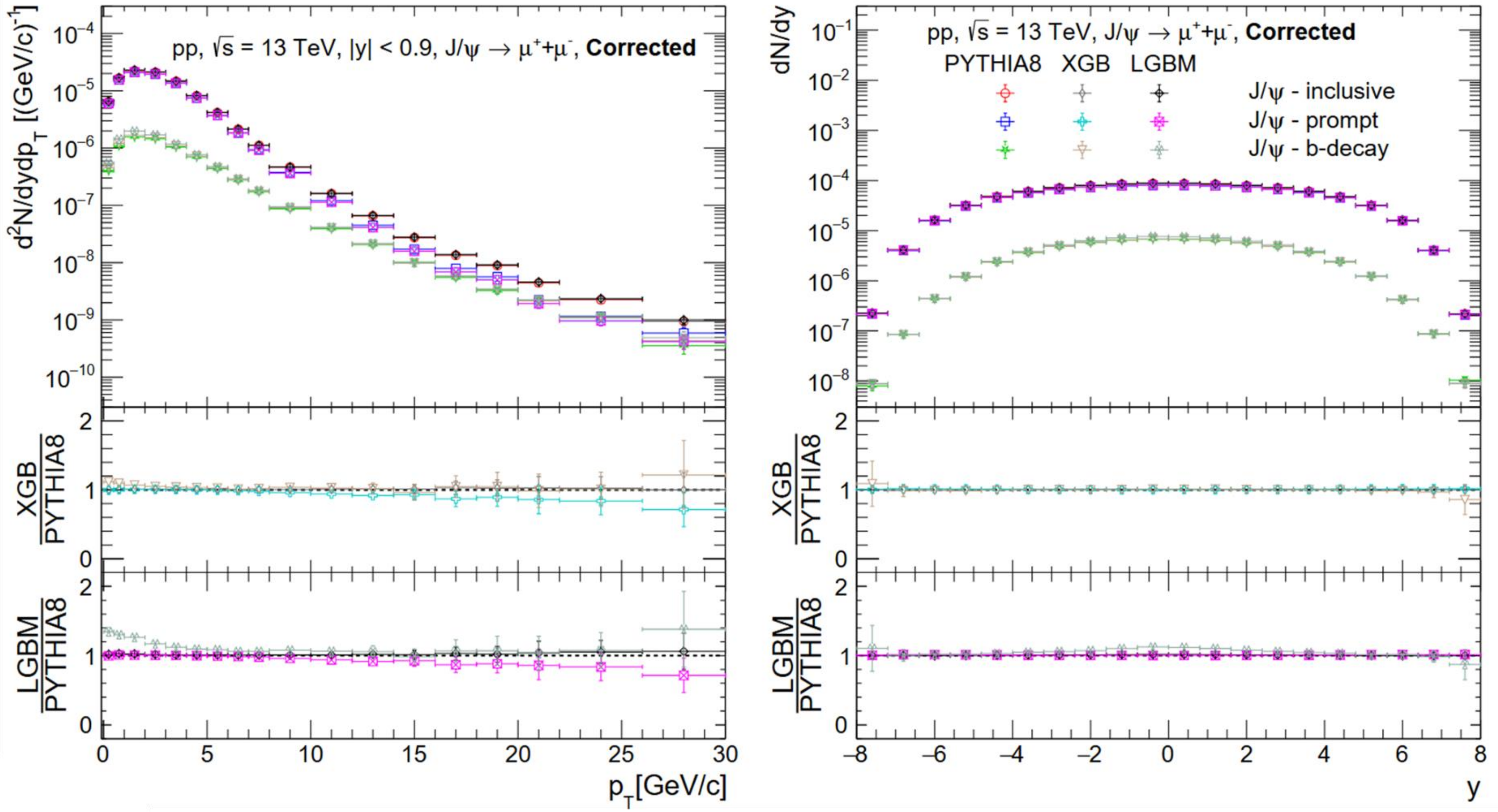
- Light Gradient Boosting Machine



Leaf-wise tree growth

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

Corrections in the Predictions



$$Y_{\mathrm{p,i}}^{\mathrm{corr}} = \frac{Y_{\mathrm{p,i}}^{\mathrm{uncorr}}}{1-\mathrm{f}}$$

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

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