UNIVERSITÄT SIEGEN

CPPS Center for Particle Physics Siegen

16th February 2024

Landhaus Nordhelle, Meinerzhagen

CPPS group retreat

Data analysis and fits summary

Diptaparna Biswas, Eleonora Guido, Gilberto Tetlalmatzi-Xolocotzi

Introduction

We have identified **3 main topics of interest** for the whole CPPS group:

1. **Understanding your data and motivations**
2. **Fit techniques**
3. **Machine Learning in Physics**

For each one of the 3 points we have:

- introduced the topic and its basic concepts through some examples from our experience and/or our group work
- suggested useful tools and techniques
- left some time *for discussion* and for comments

Outline

1. **Understanding your data (6 mins)**
    1. Introduction (1 min) - Eleonora
    2. Examples: data analyses in astroparticle physics (5 mins) - Eleonora

2. Discussion (4 mins)

3. **Fitting methods (35 mins)**
    1. Non-leptonic amplitudes from fits to data (20 mins) - Gilberto
    2. Signal+Background fits using PDFs and HistFactory (5 mins) - Diptaparna
    3. Profile likelihood unfolding (5 mins) - Buddha
    4. EFTfitter: Interpreting measurements in the context of EFT (5 mins) - Jan

4. Discussion (10 mins)

5. **Machine learning (20 mins)**
    1. CNNs applied to a photon search in Auger (10 mins) - Eleonora
    2. BDT and hyperparameter optimization (5 mins) - Arpan
    3. Jet flavour tagging using Deep Sets and GNN (5 mins) - Diptaparna

6. Discussion (15 mins)

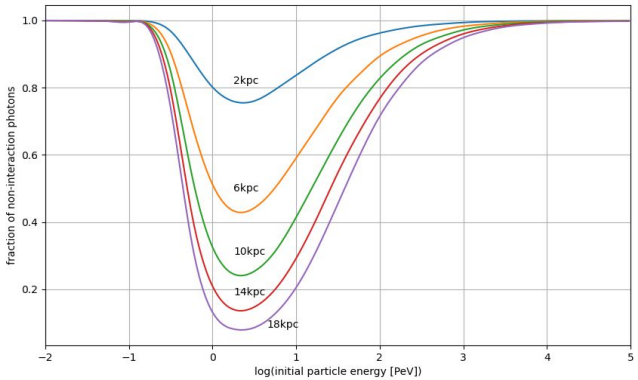
Understanding your data through data visualization

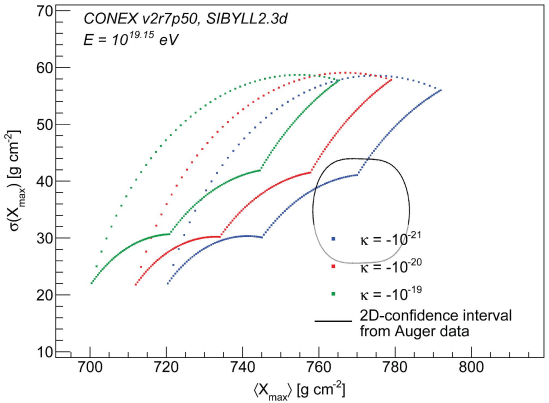
- Data data visualization is the representation of data in graphical/visual format

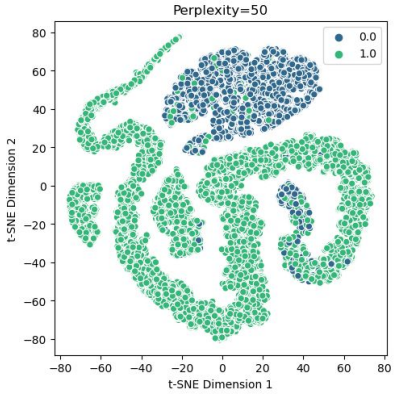
- **We highlighted its importance for:**
  - Enhancing the *understanding of physical phenomena*
  - Helping in *identifying patterns*, *trends* and which variables are the most significant
  - Effectively *communicating the results* to any audience

- We suggested common tools for data visualization, for example the **Python libraries matplotlib and seaborn**

Examples of data visualization for delivering results



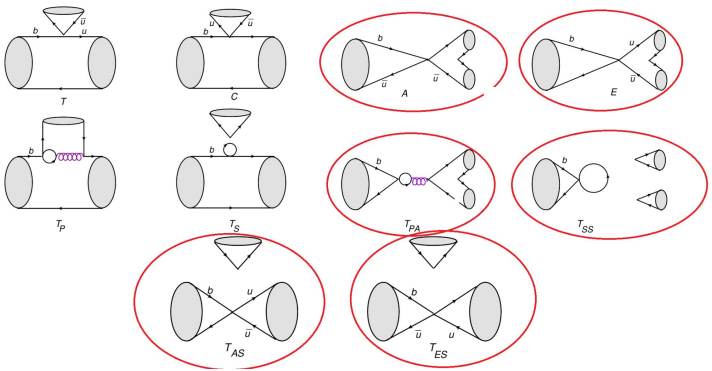


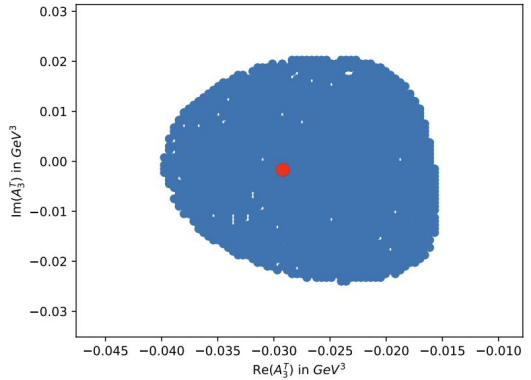


- Evaluating the interactions affecting photons during their propagation (simulated with CRPropa) . Tools: Pyplot and numpy.histogram
- Testing the impact of LV on the development of simulated air showers. Tools: ROOT
- Visualizing high-dimensional data sets. Tools: scikit-learn and seaborn

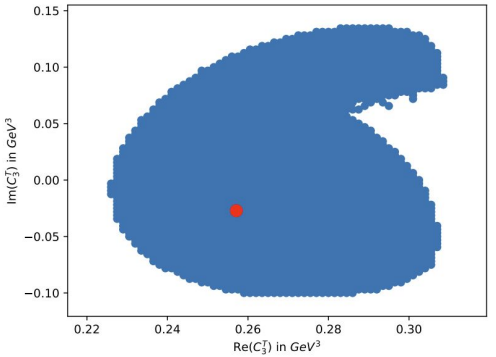
Chiara Papior, Fabian Dünkel, Eleonora Guido

χ^2 fit: Non-leptonic amplitudes from fits

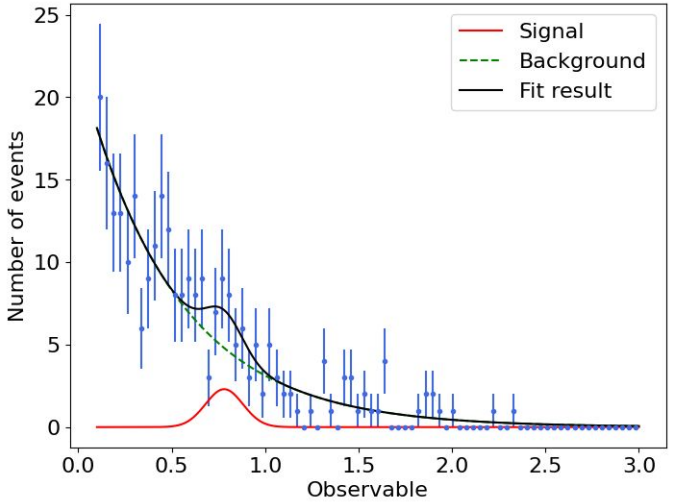




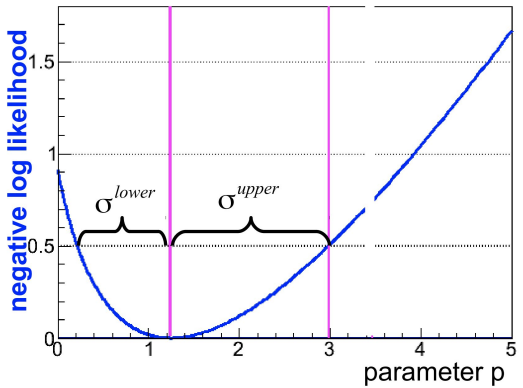
$$B^- \to \pi^0\pi^- \quad B^- \to K^0K^- \quad B^0 \to \pi^+\pi^-$$
$$B^0 \to \pi^0\pi^0 \quad B^0 \to K^+K^- \quad B^0 \to K^0\bar{K}^0$$
$$B_s \to \pi^0K^0 \quad B_s \to \pi^-K^+ \quad B^- \to \pi^0K^-$$

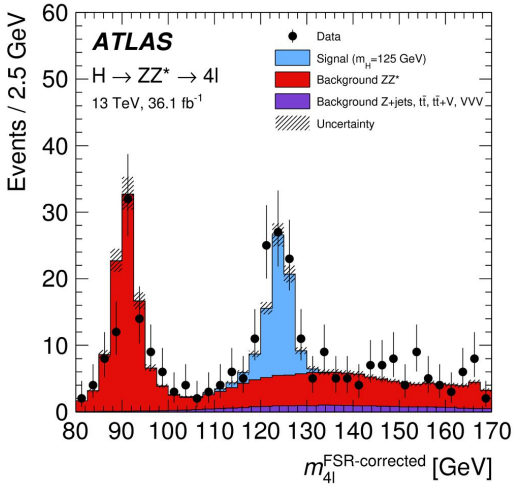


Signal+Bkg. fits using smooth PDFs and HistFactory



$$L(\vec{p}) = \prod_{n=0}^{N} F(\vec{p}, \vec{x}_n) \cdot \underbrace{Poisson(N_{exp}, N_{obs})}$$

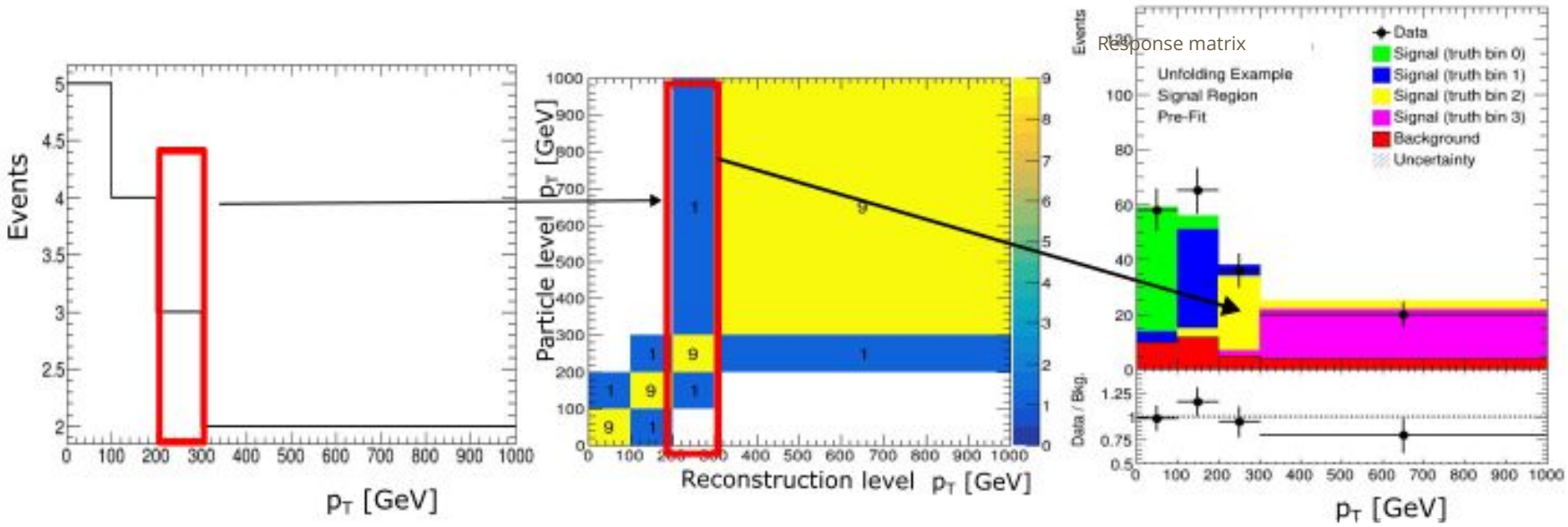




$$\mathcal{P}(n_c, x_e, a_p \mid \phi_p, \alpha_p, \gamma_b) = \prod_{c \in \text{channels}} \left[\text{Pois}(n_c|\nu_c) \prod_{e=1}^{n_c} f_c(x_e|\boldsymbol{\alpha}) \right] \cdot G(L_0|\lambda, \Delta_L) \cdot \prod_{p \in \mathbb{S}+\boldsymbol{\Gamma}} f_p(a_p|\alpha_p)$$

Diptaparna Biswas

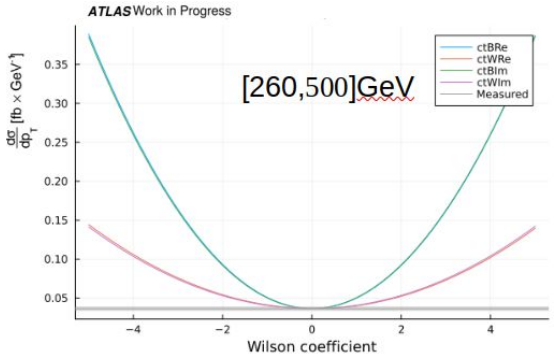
Profile likelihood unfolding

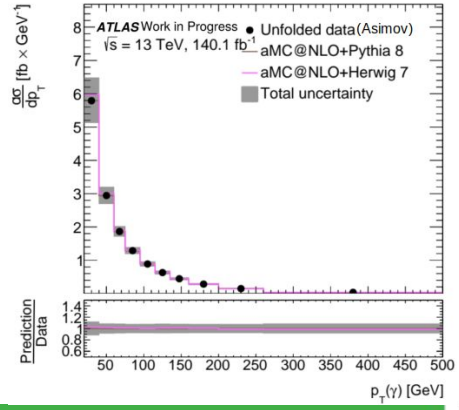


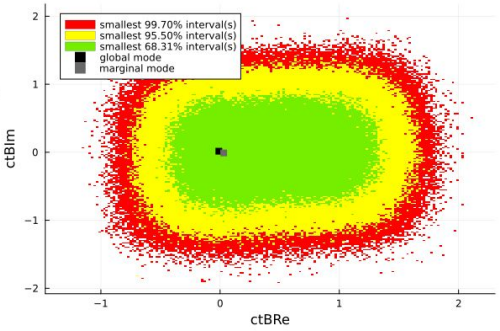
Buddhadeb Mondal

EFTfitter: Interpreting measurements in the context of effective field theories

Jan-Joachim Hahn







Measured distribution

Uncertainties

Correlations

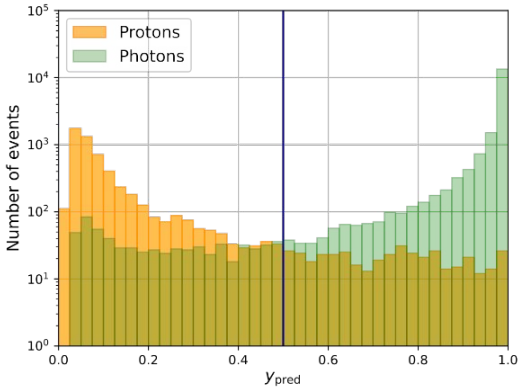
Monte Carlo EFT samples

Extract EFT dependence

EFTfitter

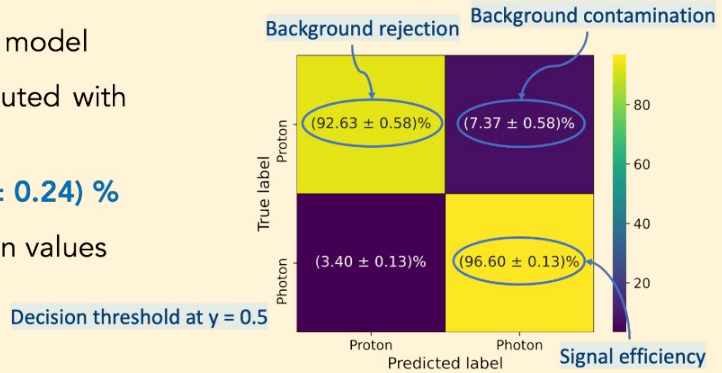
Limits

CNN applied to a UHE photon search



- Metrics are evaluated for each model
- The average results are computed with their uncertainty:

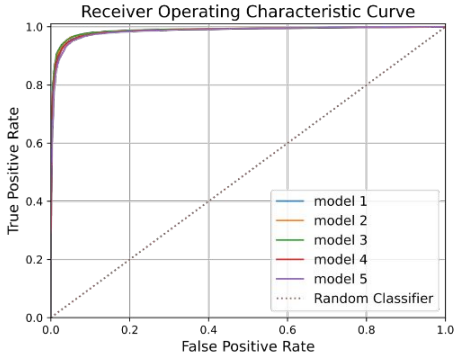
  ➜ **Balanced accuracy: (94.62 ± 0.24) %**
- Confusion matrix with the mean values

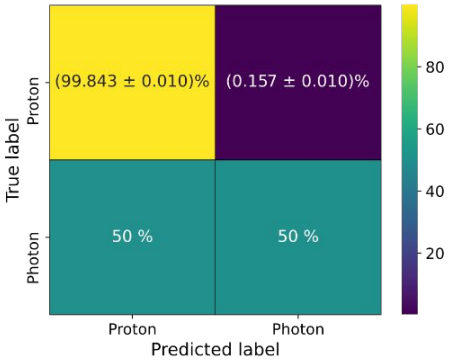


Decision threshold at y = 0.5

- In order to compare our results with other Auger analysis we are interested in **the background rejection at 50% signal efficiency**
- The decision threshold y = 0.99 provides a 50% signal efficiency

  ➜ **Background rejection is (99.843 ± 0.010)%**
- Similar to Auger analyses <u>with hybrid events</u> (99.1-99.9 %)





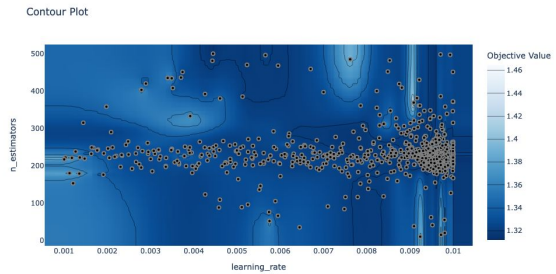
Eleonora Guido

Signal/Background separation using BDT and hyperparameter optimization

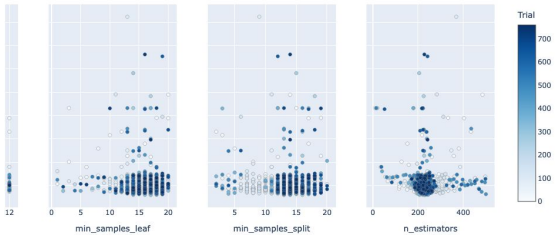
- Study results can be stored as structured database.
- Plotly - helps with visual plots of Optuna's search space.

```
study = optuna.create_study(study_name='trial1',load_if_exists=True,storage="sqlite:///./results.db",
direction="minimize")  # minimizing loss here
```

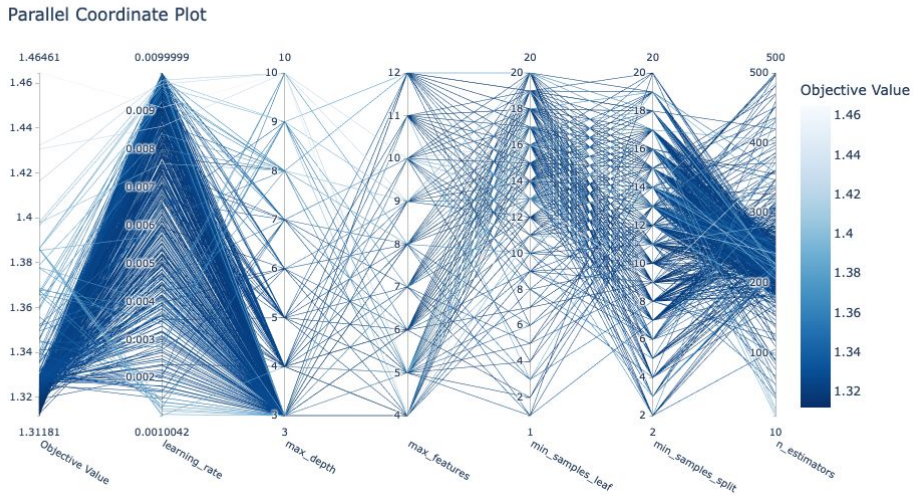
```
from optuna.visualization import plot_contour

plot_contour(study, params=['learning_rate', 'n_estimators'])
```




```
from optuna.visualization import plot_slice

plot_slice(study)
```

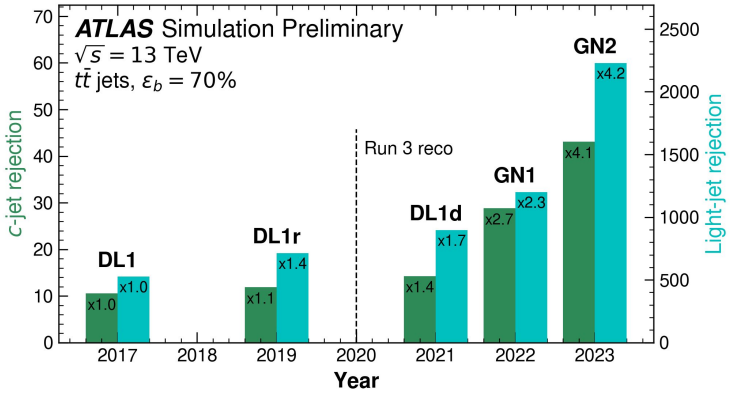


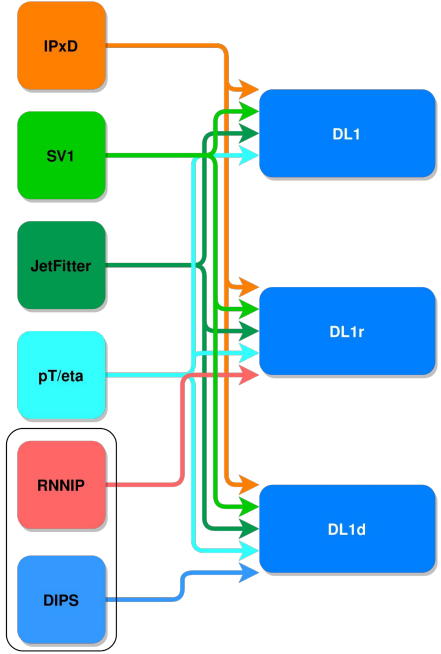

```
from optuna.visualization import plot_parallel_coordinate

plot_parallel_coordinate(study)
```



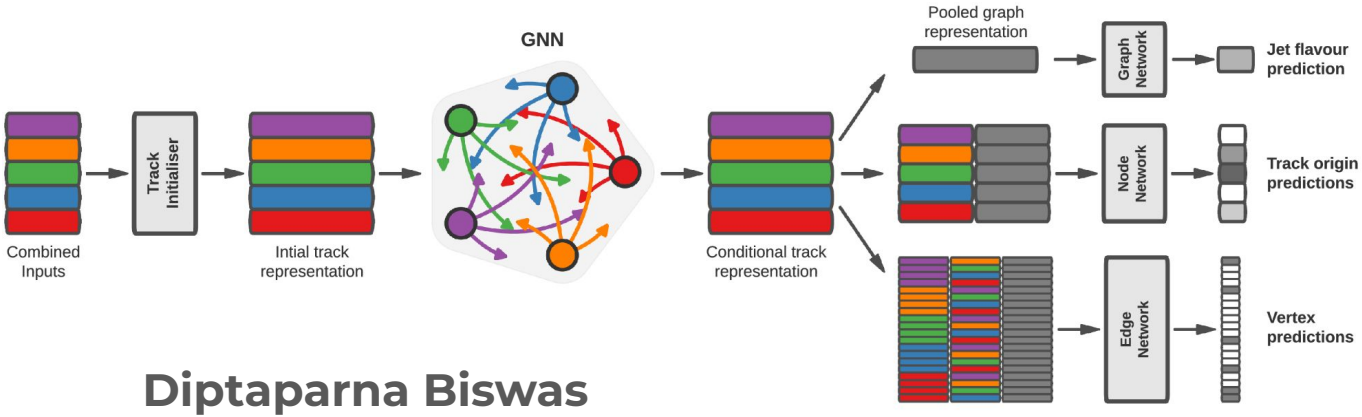
Arpan Ghosal

Jet flavour tagging using Deep Sets and GNN





Track-NN based



Diptaparna Biswas