Interpretable machine learning approach for electron antineutrino selection in the JUNO experiment

Arsenii Gavrikov[a] on behalf of the JUNO Collaboration
University & INFN Padova, Italy; arsenii.gavrikov@pd.infn.it

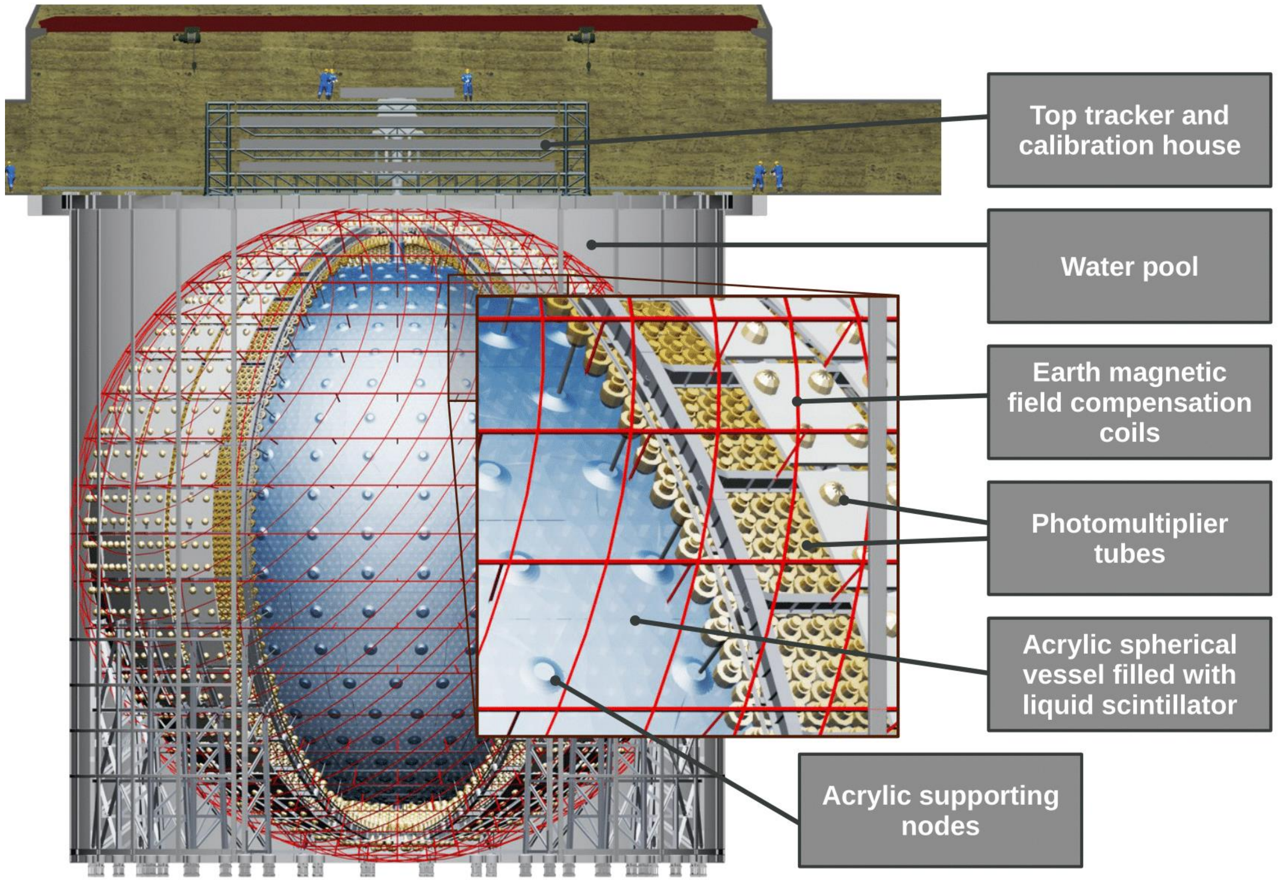
25/06/2024, Zürich, Switzerland
Neutrino Physics and Machine Learning Workshop 2024

Introduction

Introduction: the JUNO experiment

- Jiangmen Underground Neutrino Observatory (**JUNO**) [1]:
  - **multi-purpose neutrino experiment** located in China
  - **the largest 20 kt liquid scintillator** (LS) detector
  - ~35 m of diameter
  - 53 km away from 8 reactor cores
  - ~650-meters deep underground
  - ~78% photo-coverage by photo-multiplier tubes
  - **at the latest stage of its construction**

- The main goals of JUNO [2, 3]:
  - **neutrino mass ordering** with **3σ** in ~6-7 years of data-taking
  - **sub-percent measurements** of the following **oscillation parameters:** $\sin^2\theta_{12}, \Delta m^2_{21}, \Delta m^2_{31}$



[1] JUNO Collaboration 2016 *J. Phys. G: Nucl. Part. Phys.* **43** 030401
[2] JUNO Collaboration 2022 *Chinese Phys. C* **46** 123001
[3] JUNO Collaboration 2024 *arXiv:2405.18008*

Introduction: neutrino detection in JUNO

- "Golden" channel to detect the reactor antineutrinos in a LS-based detector: **Inverse Beta Decay (IBD)**

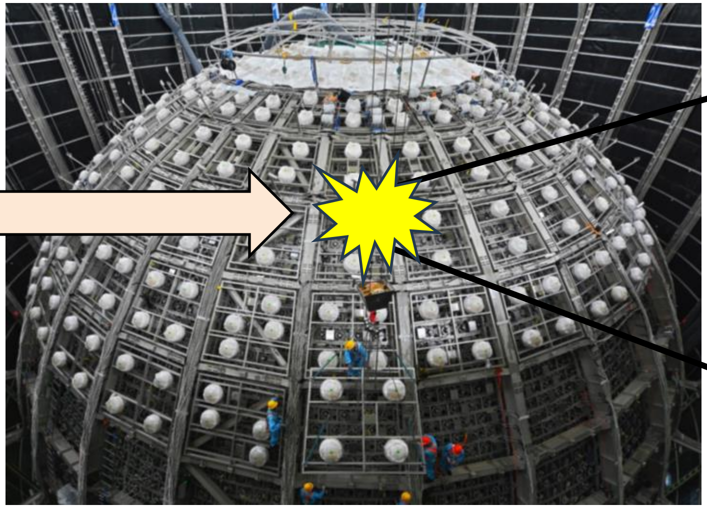
$$\overline{\nu}_e + p \rightarrow e^+ + n$$

- **e^+** takes most of the neutrino energy and annihilates with an **e^-**, forming **a prompt signal**
- **the neutron** thermalizes and **after ~220 μs gets captured** on an isotope of the target medium: H (~99% cases), C (~1% cases)
  - subsequent emission of **2.22 MeV (H)** or **4.95 MeV (C) gammas** generates **a delayed signal**
- This **double signature** represents a powerful tool to **discriminate signal from backgrounds**

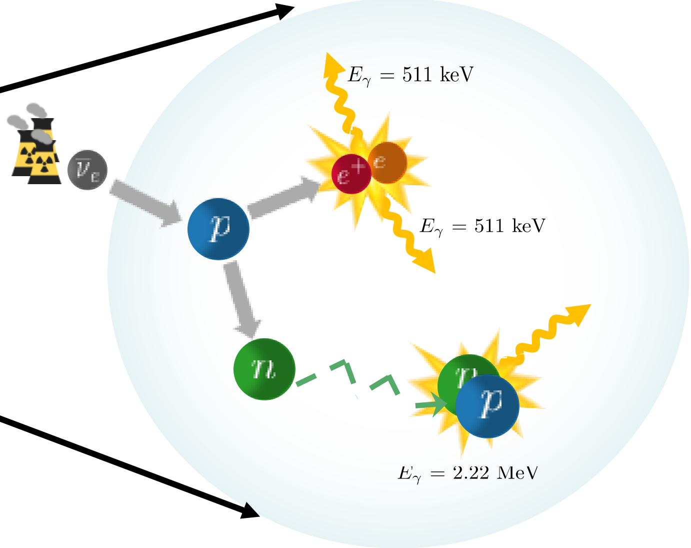
Nuclear power plant: $\overline{\nu}_e$ generator

$\overline{\nu}_e$

target volume filled with LS



IBD reaction



Data and non-ML selection

Signal and backgrounds

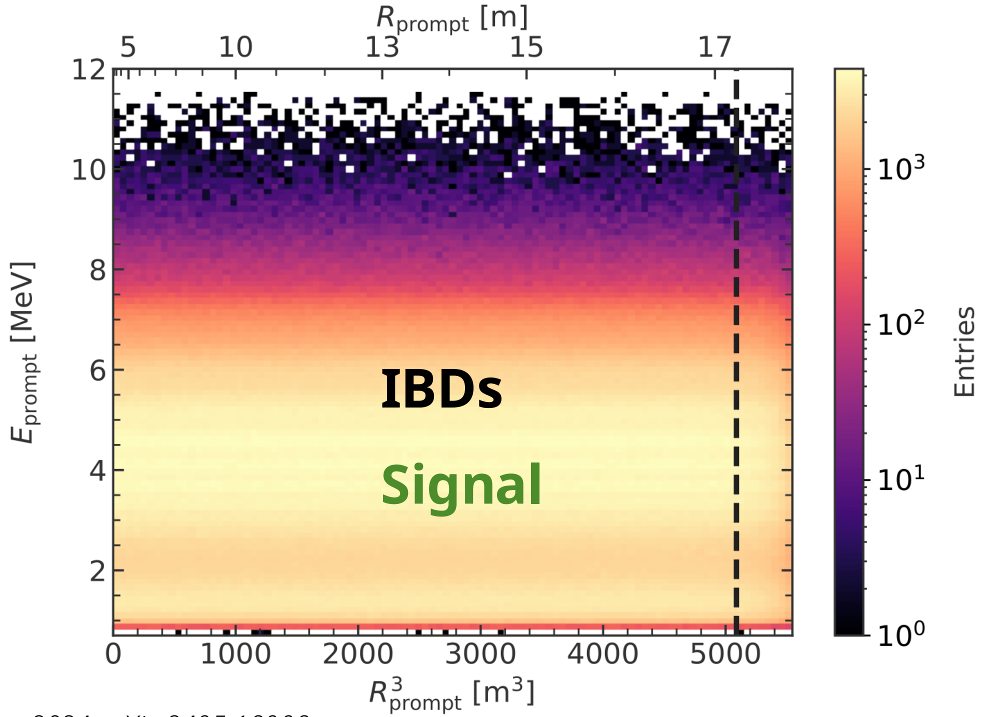
- **Signal**: reactor antineutrino events
- **Backgrounds** are primarily attributed to **radioactive contamination** in the detector materials and surroundings and **can be divided in two categories:**
  - **Uncorrelated** (accidental coincidences):
    - **two independent signals** in a short time window **mimic the IBD pattern**
  - **Correlated background:** consists of a pair induced by a single physics process, mimicking the prompt-delayed pattern (e.g. geoneutrinos, cosmogenic isotopes):
    - some correlated backgrounds are irreducible
    - others can be reduced through a muon cut but their residual contamination is considered irreducible

The main task required of **a selection algorithm** is **to distinguish between two classes**:
reactor antineutrino events and accidental coincidences.

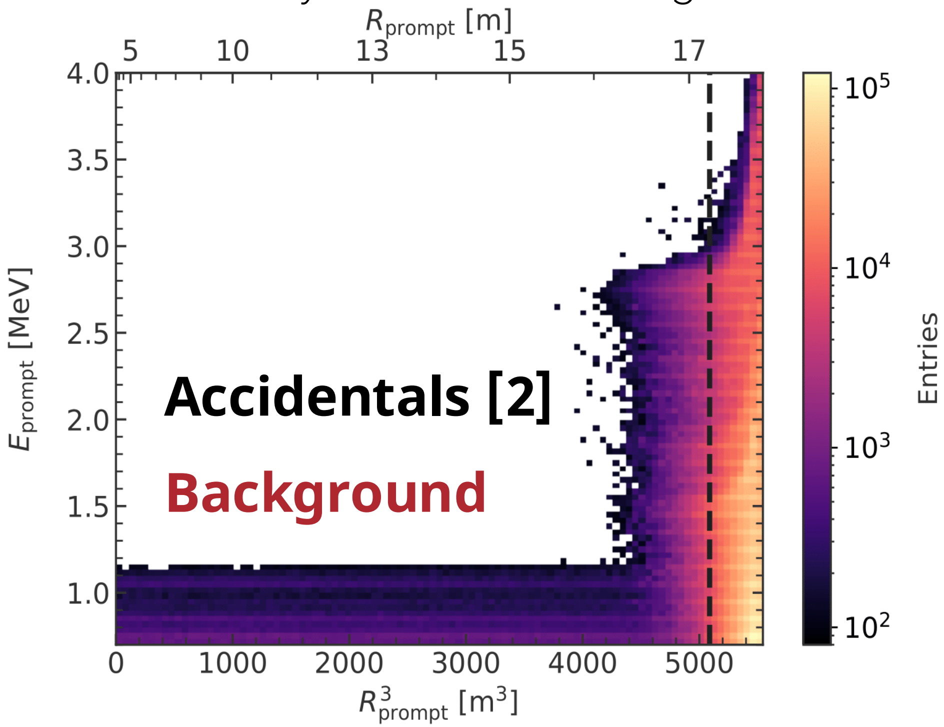
Data description

- Highly imbalanced data: 57.4 **IBDs [1]** vs. $>10^5$ **accidentals [2]** [events/day]
- For training and hyperparameter optimization, **balanced dataset is used** to prevent biases:
  - 15M of IBD pairs
  - 15M of radioactive coincidences (accidentals)
- To properly evaluate a method, **we compare it with a cut-based selection**

Uniformly distributed in the detector



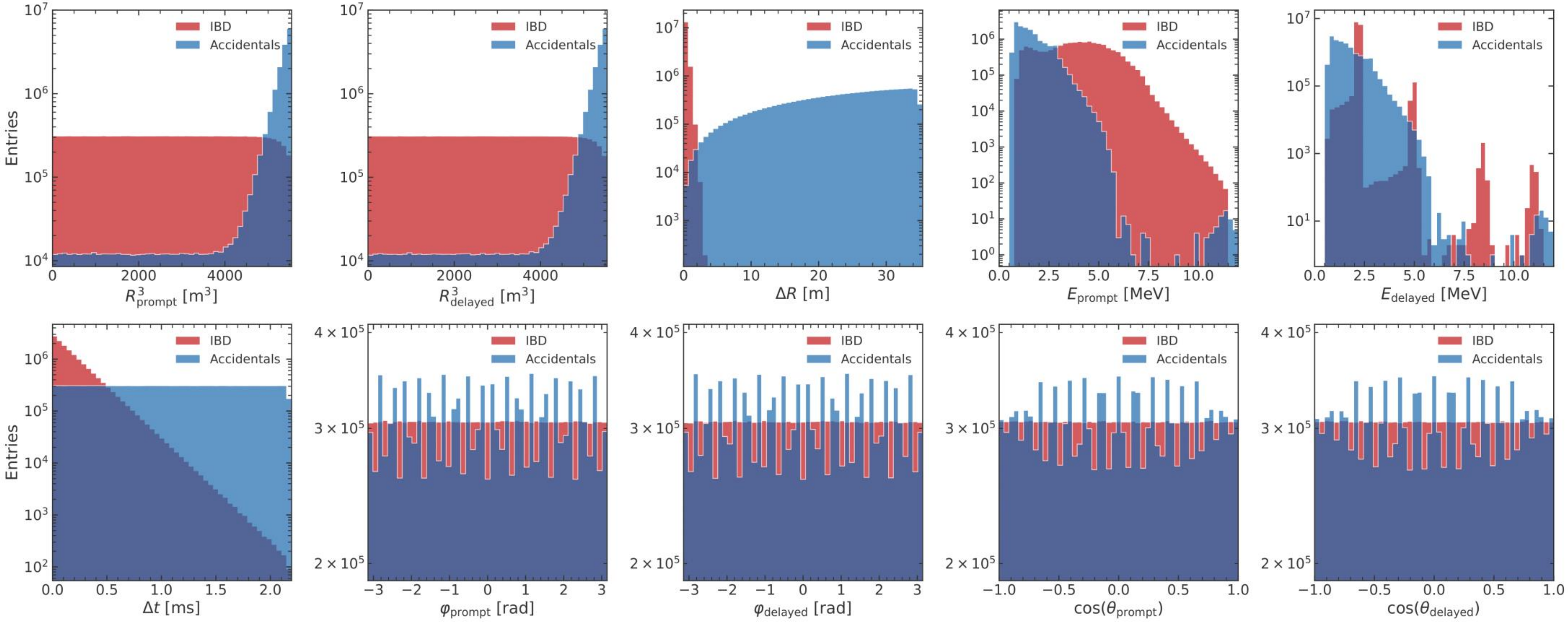
Mostly at the detector edge

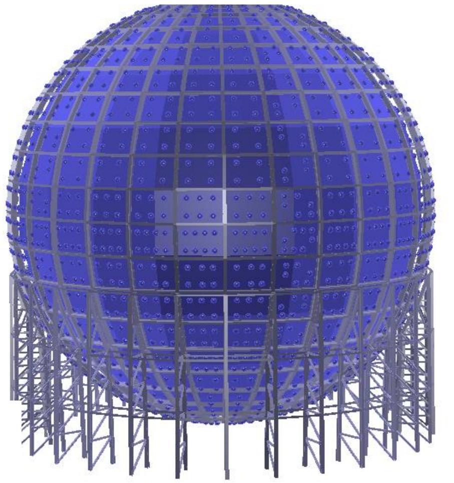


[1] JUNO Collaboration 2024 *arXiv:2405.18008*
[2] JUNO Collaboration 2021 *J.High Energ. Phys.* 102

Data description: feature distributions

10 features energy and vertex related: R^3_p, R^3_d, ΔR, Δt, E_p, E_d, $\cos(\theta_p)$, $\cos(\theta_d)$, ϕ_p, ϕ_d



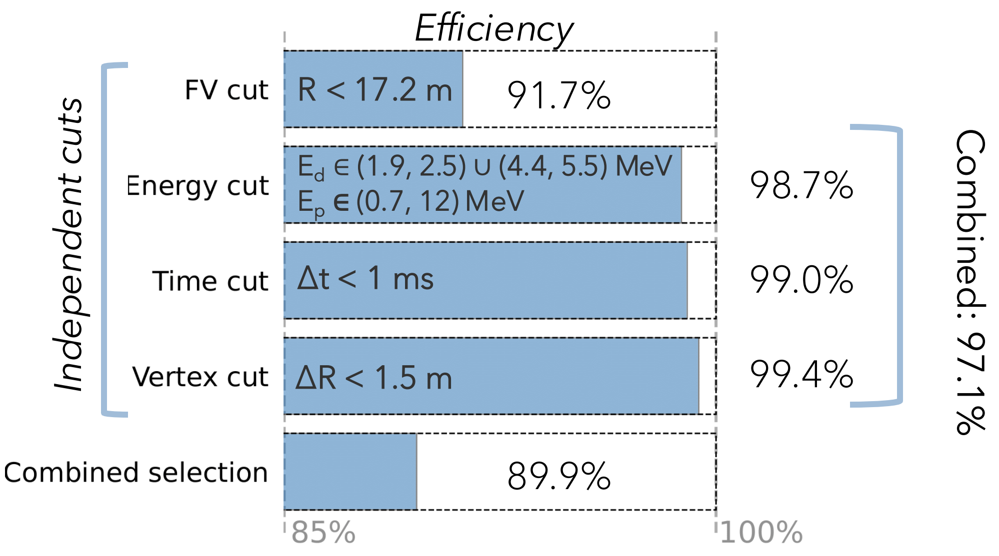


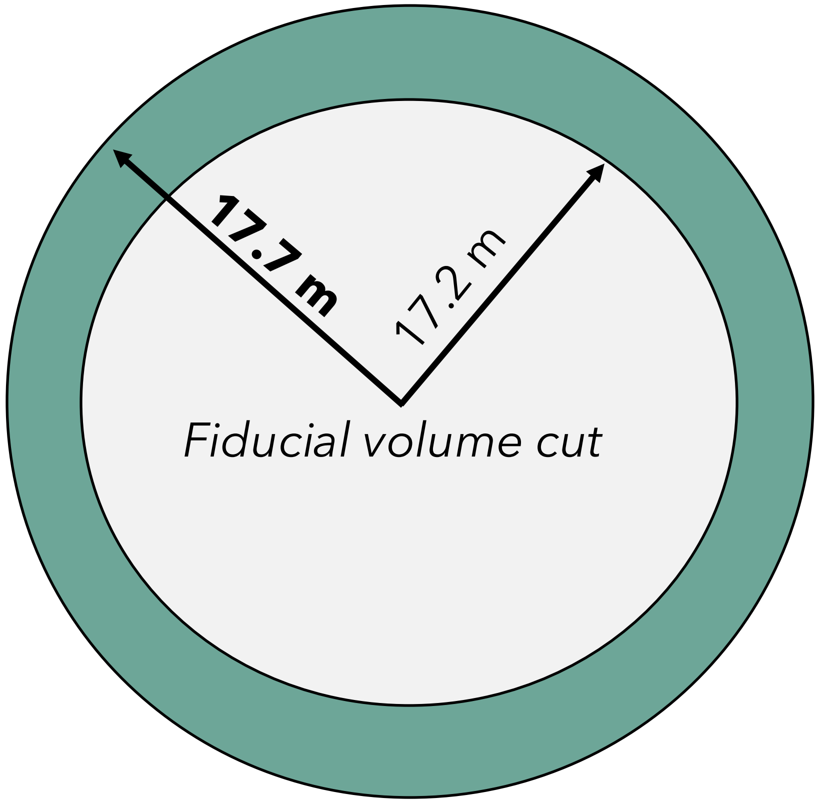
Higher radioactivity at the edges from the grid

*The dataset was produced independently of the JUNO official software

Benchmark selection: cuts

- **As a benchmark** for a machine learning approach **the cut-based selection strategy** from [1] is used
- The standard selection is effective but there is still a gap to improve:
  - Can we **remove FV cut and retain border events** saving the same level of purity? Or even **increase the efficiency within FV**?
  - Can we **improve the overall efficiency**?
- **Let's explore machine learning methods for this!**





*The cut-based selection does not necessarily reproduce JUNO official selection

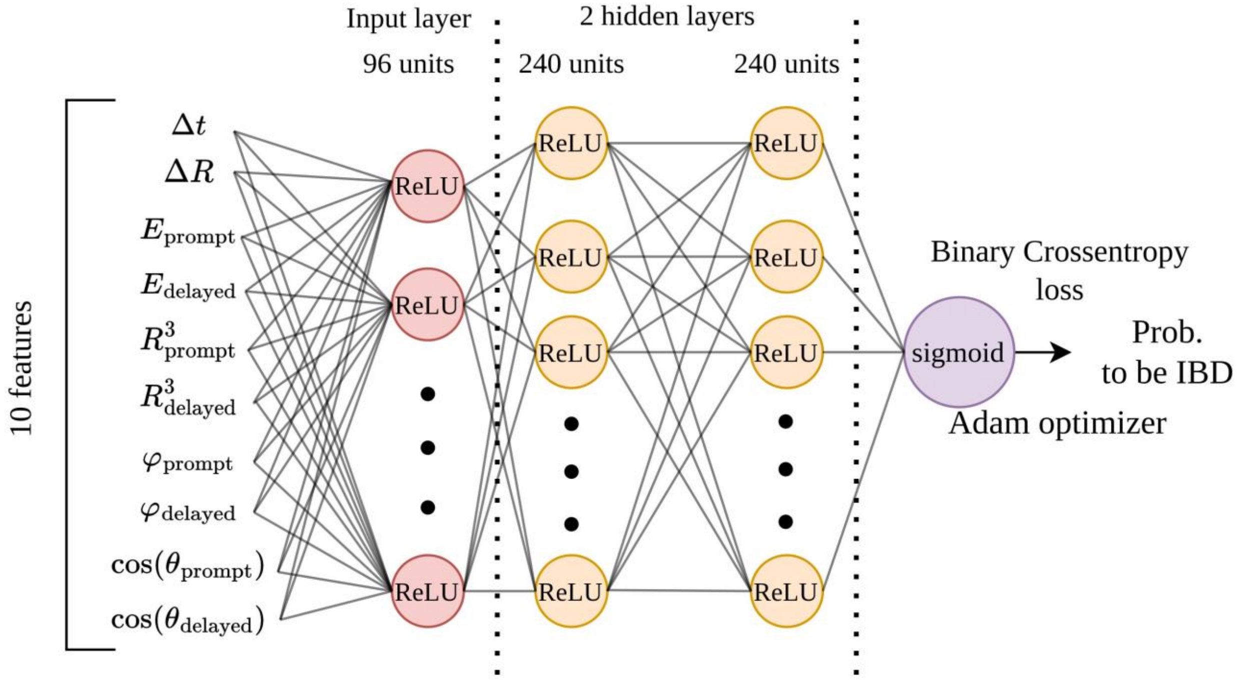
[1] JUNO Collaboration 2022 *Chinese Phys. C* **46** 123001

Machine learning approach [1]

[1] A. Gavrikov et al. 2024 *arXiv: 2406.12901*

Neural network: architecture

- 10 features as input:
  - Δt, ΔR
  - E_p, E_d
  - R^3_p, R^3_d
  - ϕ_p, ϕ_d, $\cos(\theta_p)$, $\cos(\theta_d)$
- Trained **to separate** between *IBDs* and *accidentals*
- Comprehensive **hyperparameter optimization** with Optuna library [1]
- **Binary Crossentropy** as a loss function



[1] A. Gavrikov et al. 2024 *arXiv: 2406.12901*

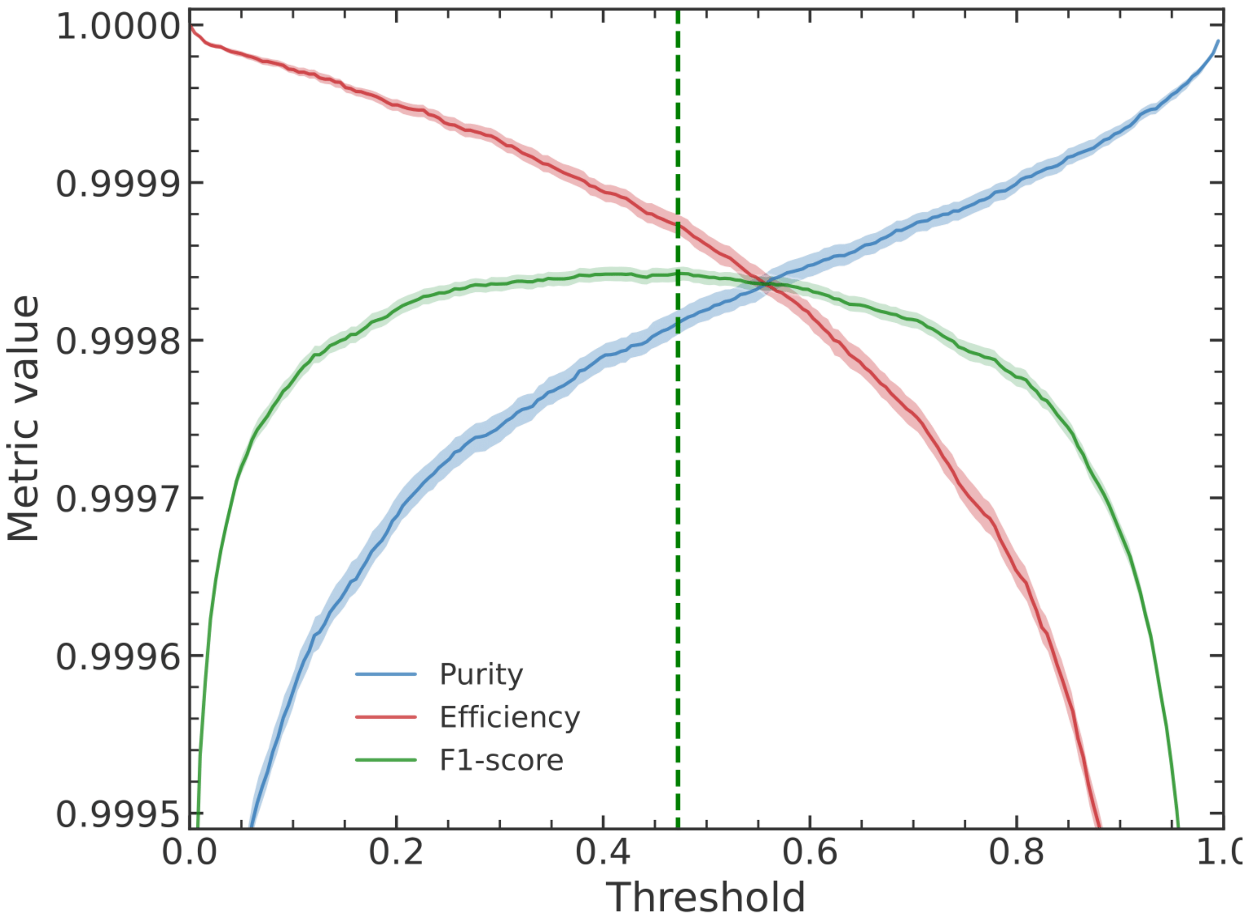
Neural network: a tunable classifier

- FCNN with ~100k weights: *fast* and *simple*
- As output: **probability to be an IBD event** from 0 to 1
- **Threshold** to assign a class is **a tunable parameter**

- For different physics channels we can use:
  - **the same model**
  - with **different thresholds**
  - **to optimize** the desired **metric** (*efficiency* or *purity*)

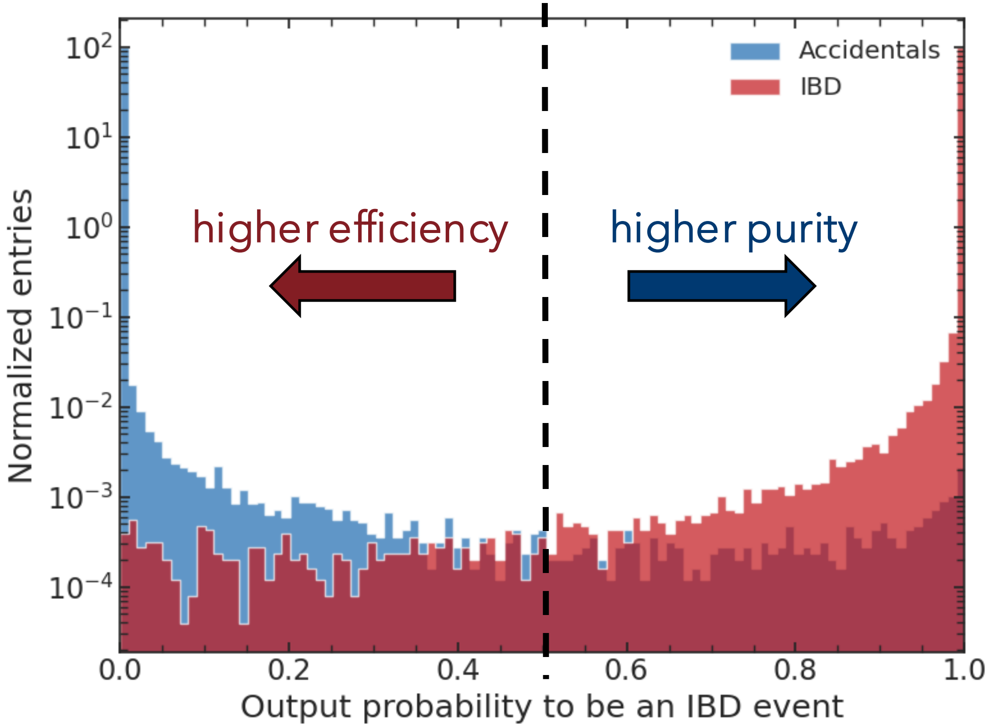
- As a baseline, we can balance them by maximizing the *harmonic mean* of efficiency and purity: **F1-score**

$$F_1\text{–score} = 2 \cdot \frac{\text{purity} \cdot \text{efficiency}}{\text{purity} + \text{efficiency}}$$

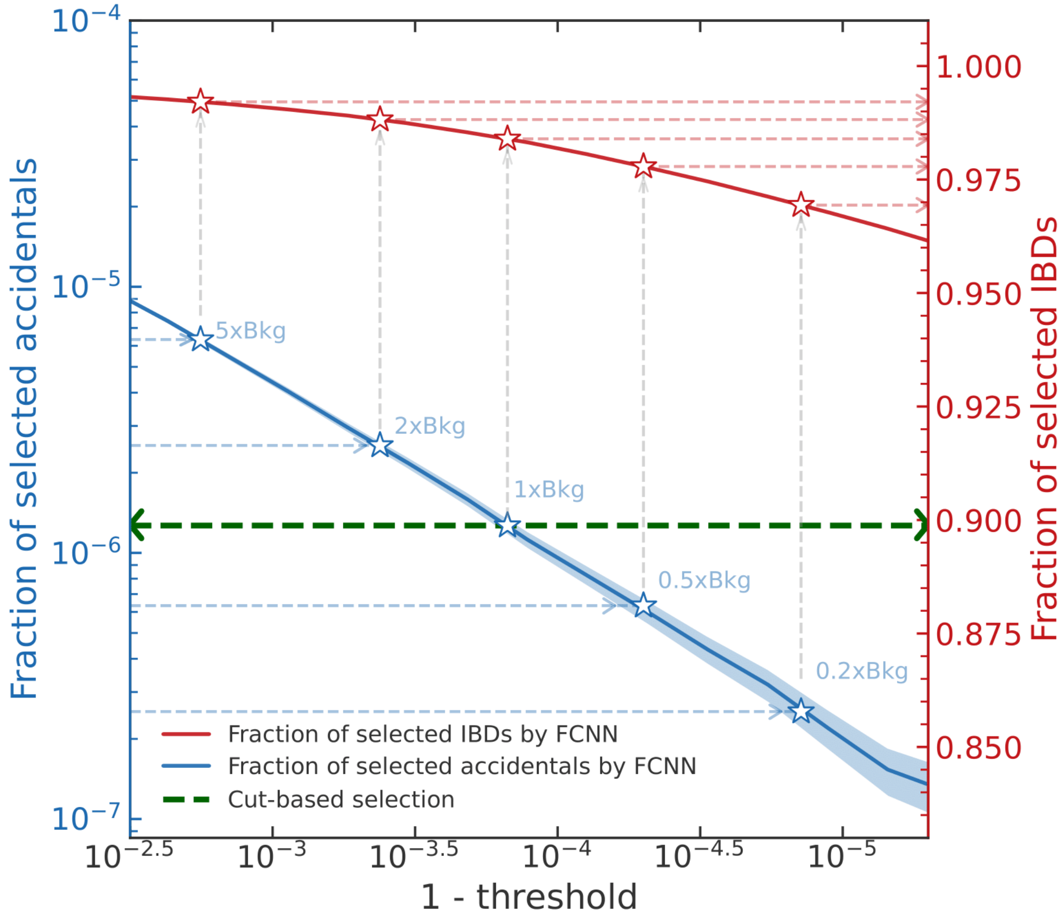


Neural network: performance

- **Tunable selection threshold** provides **flexible trade-off** between efficiency and purity
- **An improvement of ~8.5% points in efficiency** for exactly **the same background level** as for the cuts*
- *Background can be further decreased, keeping higher efficiency*
- *Efficiency can be further increased, having higher residual bkg.*



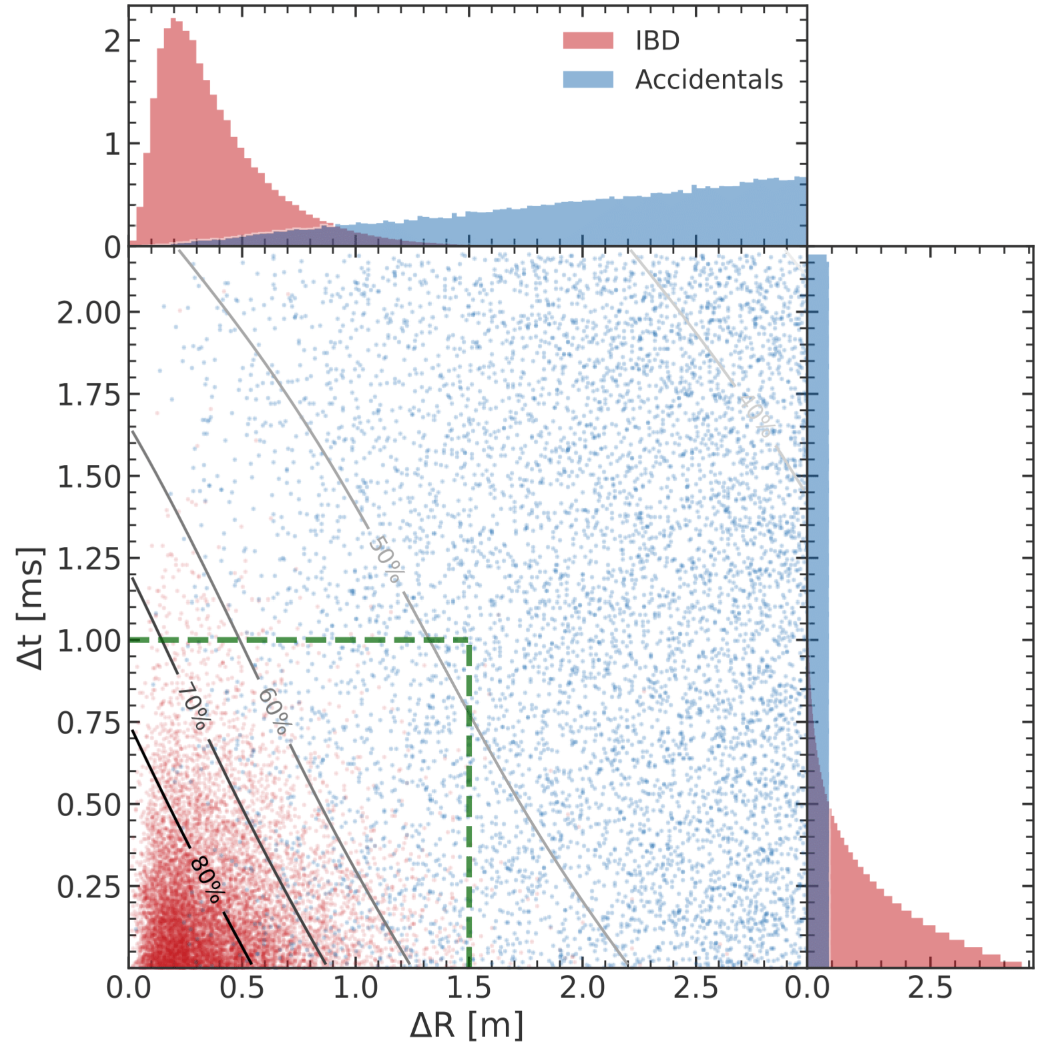
Model can be used **in the entire volume**, without fiducial volume cut



*The cut-based selection does not necessarily reproduce JUNO official selection

Neural network: optimization of IBD selection cuts

- An example of ΔR and Δt
- The box cuts (*green lines*)
  - are **effective**
  - but **can be optimized**
- The model has learned **smoother boundary**:
  - therefore improved the efficiency
  - close to linear
  - can be used as a **selection criteria**

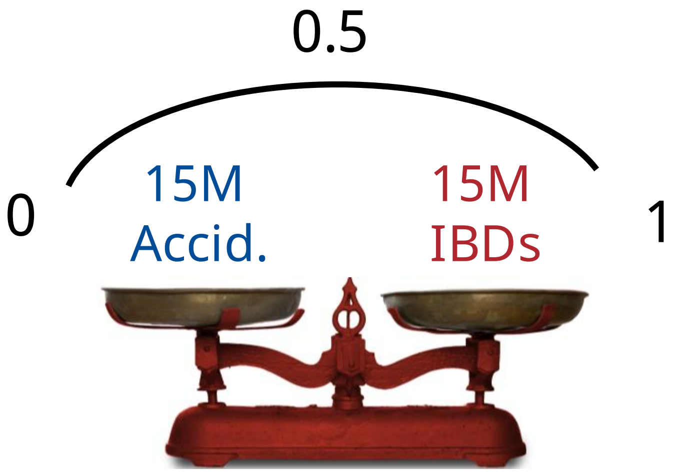


Interpretability analysis

Interpretability: "How did an ML model make a decision?"

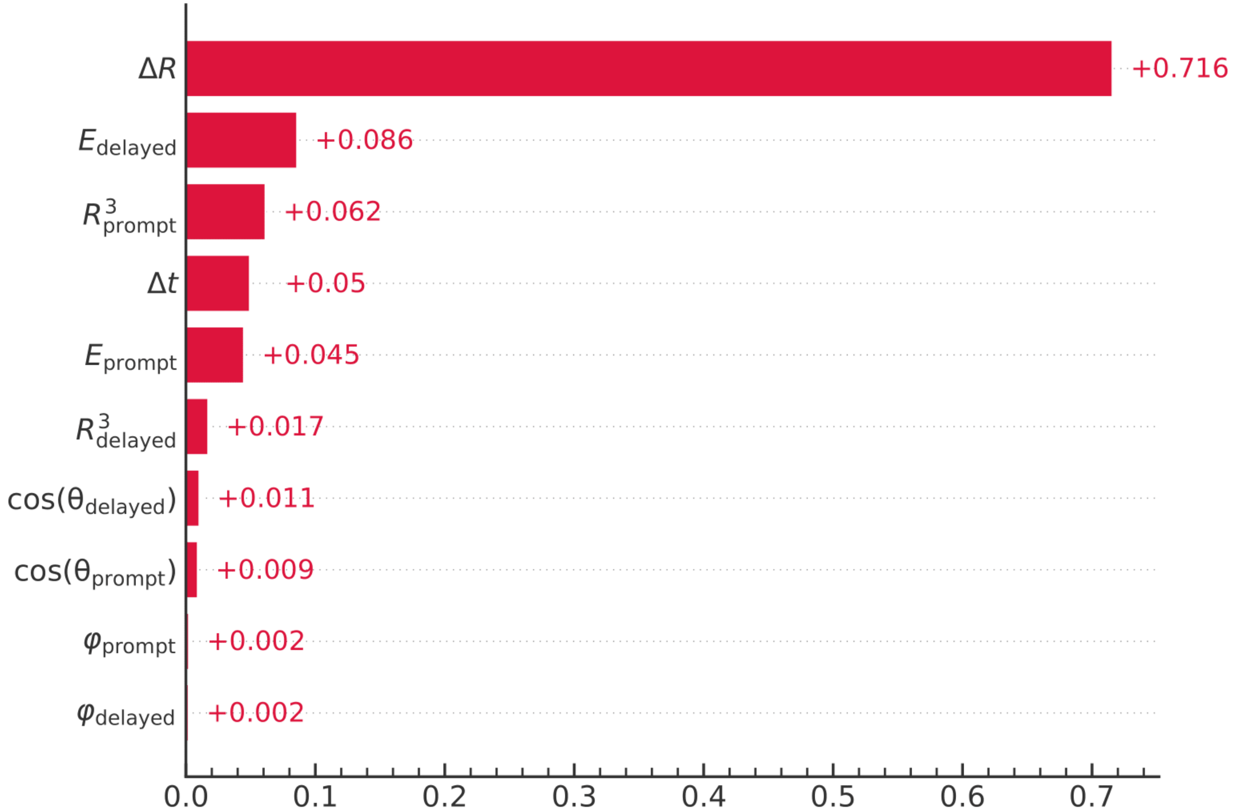
- Why explainability may be useful?
  - **Ensure trust** in the model and its **transparency**
  - **Optimize and fine-tune** *the cut-based selection strategy*

- We use SHAP (**SH**apley **A**dditive ex**P**lanations):
  - calculate the *contribution of each feature*
  - can be **both positive** and **negative**
  - show the **impact on predictions with respect to the average**:



Interpretability, *global*: "What features are most important for the model's predictions on average?"

- **Global** explanations compute the **summarized impact** of a specific feature
- Accidentals and IBD are **mostly selected on** the basis of ΔR
- **Delayed energy** has **the *next stronger*** *discriminative power*
- **Δt** is less important in general but *helps to correct* a prediction
- **R^3_{prompt}, $R^3_{delayed}$** provide additional information (especially *at the edges*)
- **Cosine θ features** behave as *a proxy to the grid*



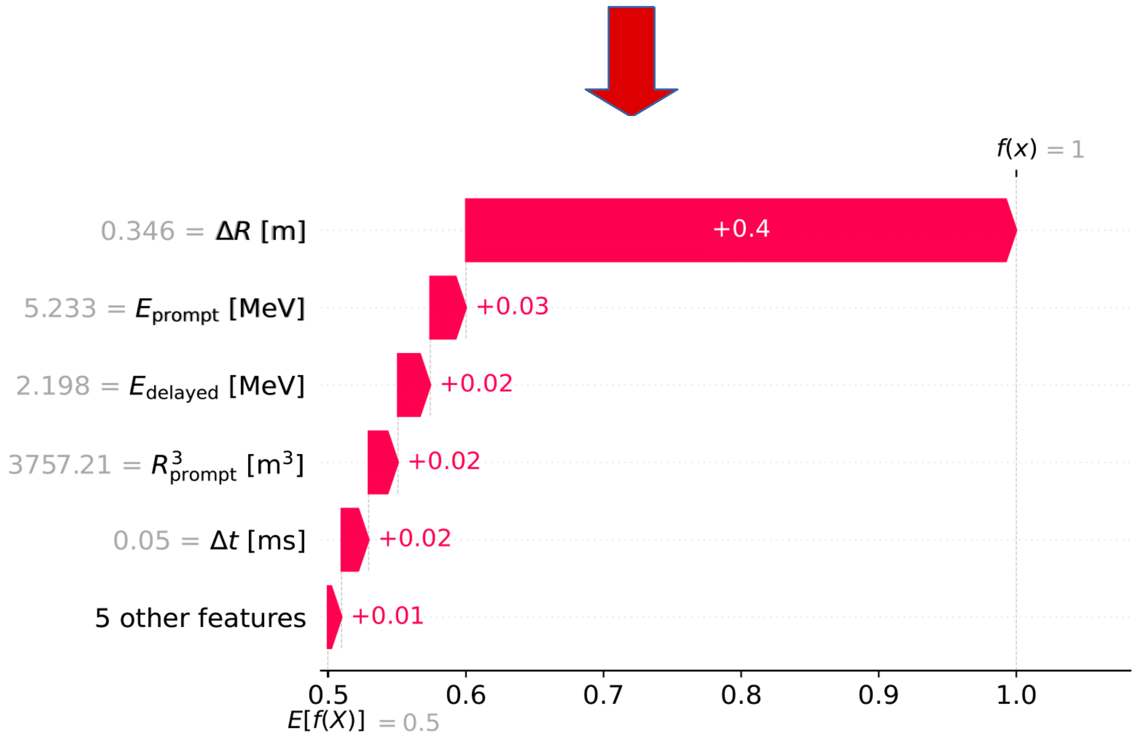
$$\phi_i = \frac{\overline{|\text{SHAP values}|}_i}{\sum_{j=1}^{10} \overline{|\text{SHAP values}|}_j}$$

Interpretability, *local*: "Why did the model make this particular prediction for this specific event?"

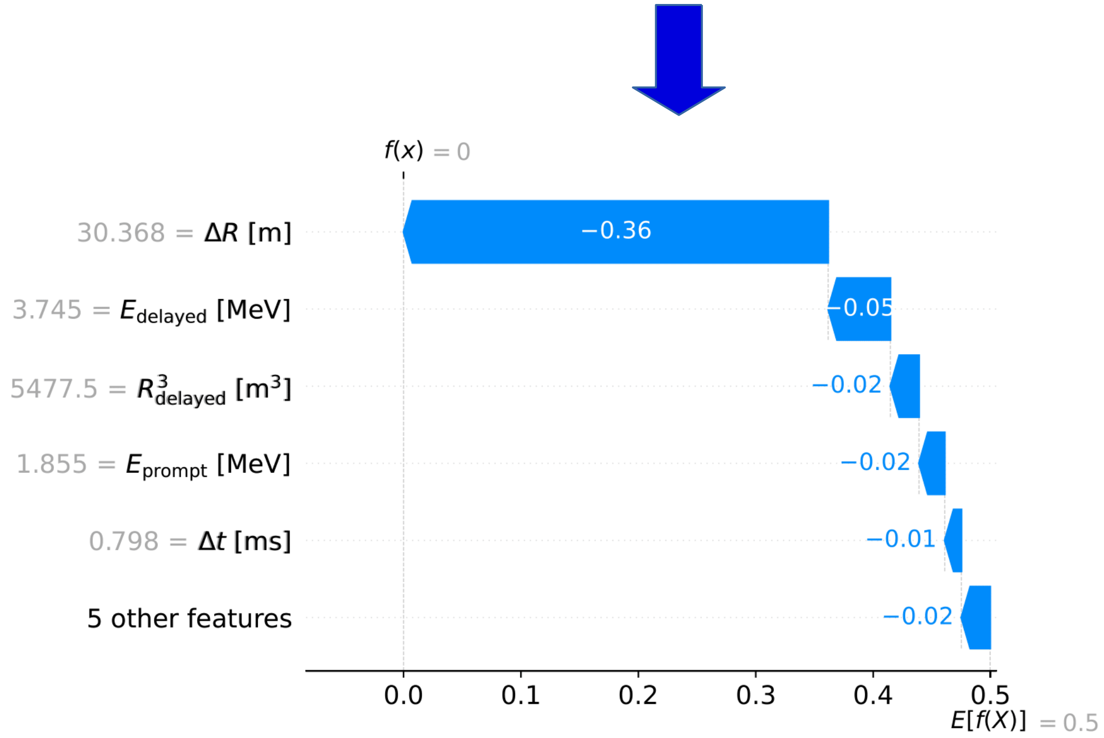
Local explanations focus on an individual event and provide features' importances for a specific instance

Typical correctly classified events:

True positive (IBDs)



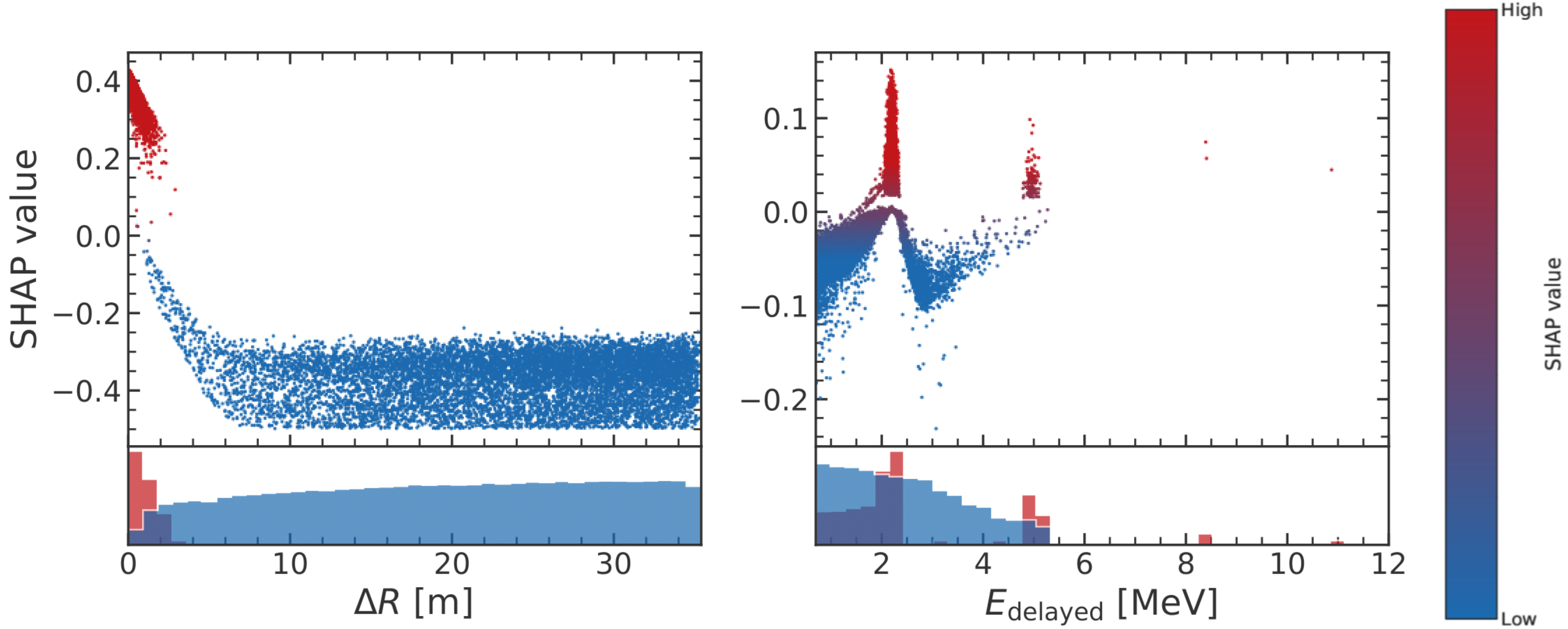
True negative (accidentals)



Interpretability: diving into the details

- Back from local explanations to **the global picture!**
- SHAP values for **top-2 features** and their distributions
- Each point is an event: 20k in total

- **As examples:** small values of ΔR are a huge indication of an IBD event
- E_d has a clear clustered structure related to specific isotope on which **neutron** was captured



Conclusions

Conclusions

- ML approach:
  - uses a fully-connected **neural network** as the IBD-accidentals classifier
  - can be applied as **a proxy to improve the cut-based selection**
  - can be used as **a stand-alone algorithm**
  - **provides higher efficiency** than the cut-based selection within both *fiducial volume* and the *entire volume*
  - keeps **higher efficiency** even with **better purity**
- Interpretability analysis of the model:
  - **ensures the robustness of model predictions** both at the local and global level
  - **offers valuable insights** into model behavior
- *More details on the approach and the results in a preprint [1]*

[1] A. Gavrikov et al. 2024 *arXiv: 2406.12901*

Thank you!

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101034319 and from the European Union – NextGenerationEU.