INTERPETABLE MACHINE LEARNING APPROACH FOR ELECTRON ANTINEUTRINO SELECTION

ARSENII GAVRIKOV, **VANESSA CERRONE**, ANDREA SERAFINI, et al.

based on Phys. Lett. B 860 (2025) 139141

vanessa.cerrone@studenti.unipd.it

Applied Antineutrino Physics 2024 workshop
Aachen , 28-30 October 2024

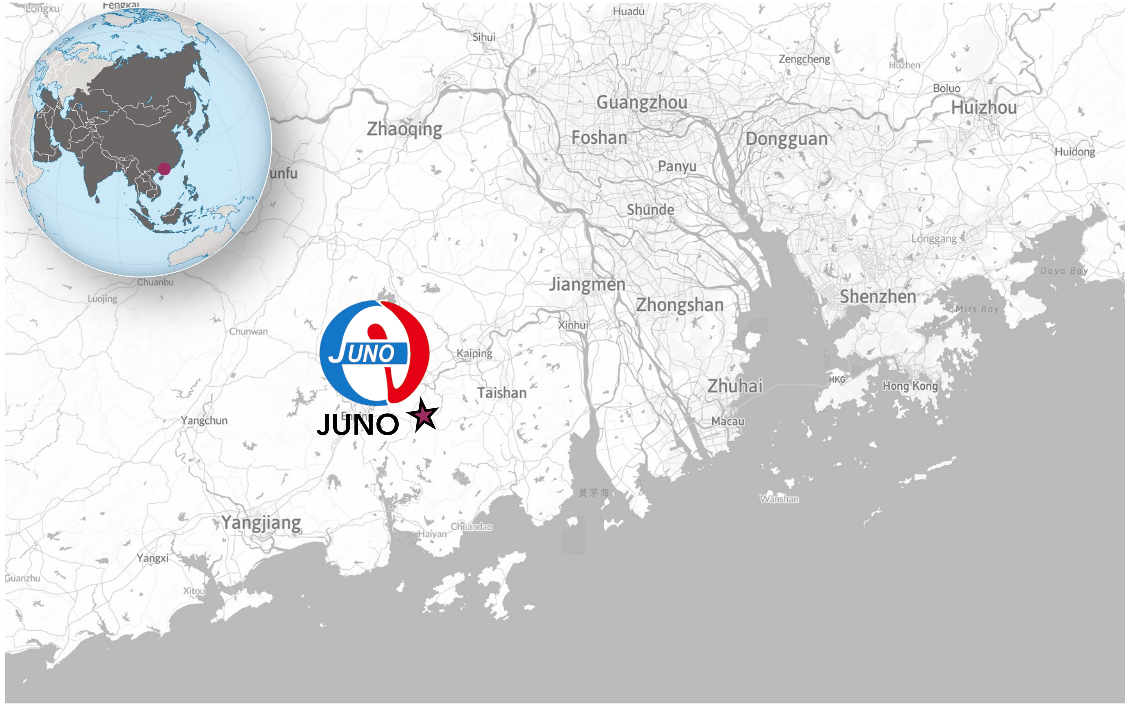
INFN Istituto Nazionale di Fisica Nucleare

Università degli Studi di Padova

JUNO AT A GLANCE

PPNP 123 (2022): 103927

The **J**iangmen **U**nderground **N**eutrino **O**bservatory (**JUNO**) is a multi-purpose neutrino experiment currently under construction in South China.

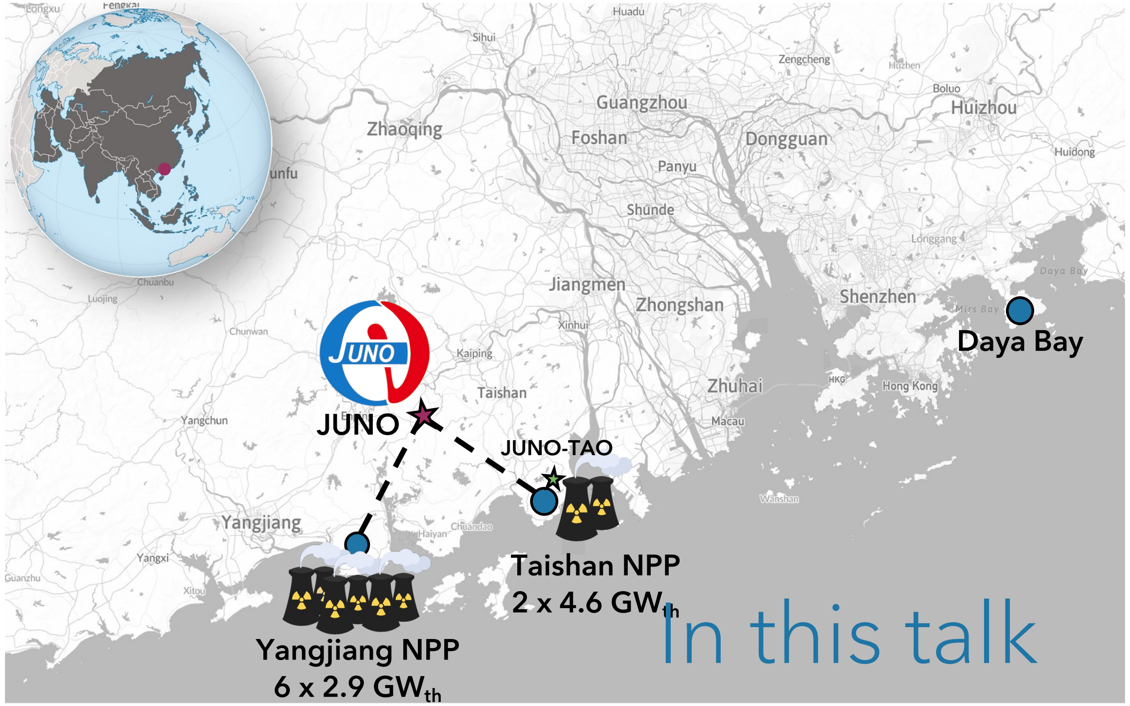


JUNO AT A GLANCE

PPNP 123 (2022): 103927

The **J**iangmen **U**nderground **N**eutrino **O**bservatory (**JUNO**) is a multi-purpose neutrino experiment currently under construction in South China.

- ✶ 52.5 km from two major Nuclear Power Plants (NPPs) with eight nuclear reactors (26.6 GW_{th})
- ✶ 35 m-diameter sphere with **20 ktons of liquid scintillator** (LS) surrounded by a water Cherenkov detector
- ✶ Unprecedented energy resolution for a LS-based detector → **3% at 1 MeV** arXiv 2405.17860



In this talk

Experiment	Daya Bay	RENO	Double Chooz	KamLAND	JUNO
LS mass	20 ton	16 ton	8 ton	1 kton	20 kton
Energy resolution	8%	8%	8%	6%	3%

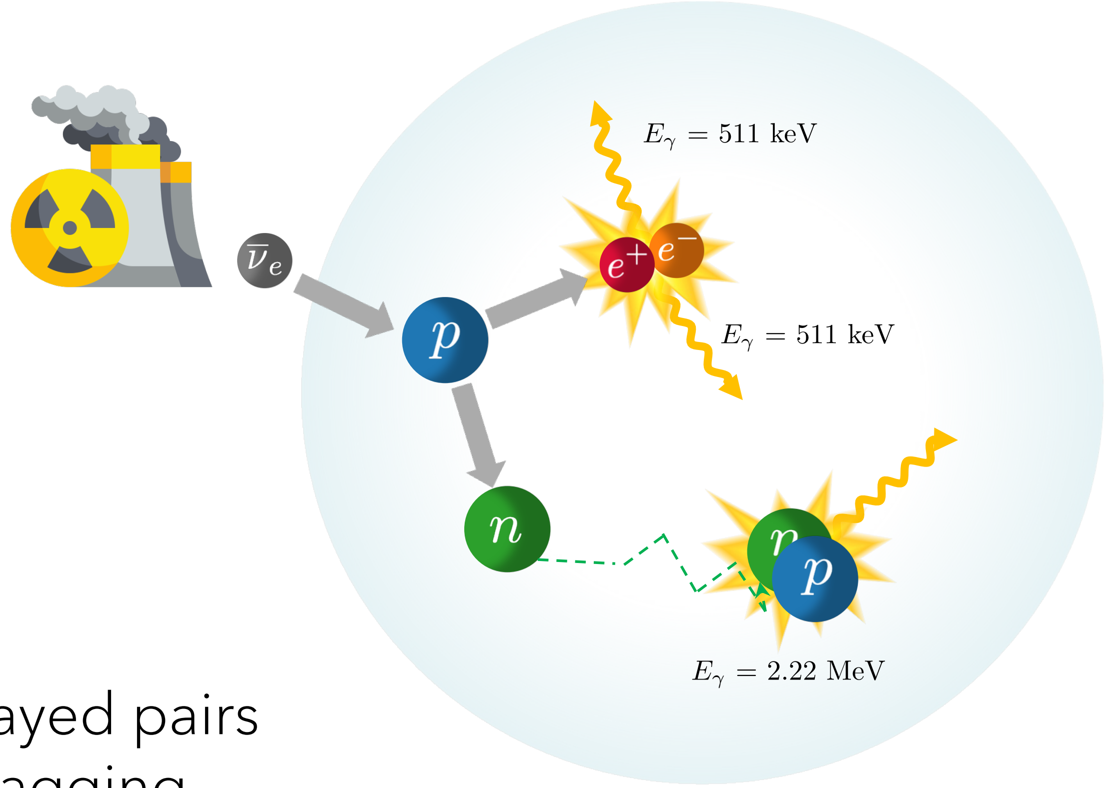
ANTINEUTRINO DETECTION

✶ Primary detection channel is the Inverse Beta Decay (IBD) reaction:

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

Prompt signal: energy deposited by positron + annihilation energy

Delayed signal: neutron capture on nucleus (e.g., Hydrogen, Lithium, Gadolinium) and subsequent gamma ray emission



✶ Close time and space correlation of prompt-delayed pairs
→ efficient background suppression and event tagging

✶ Positron retains most of incoming antineutrino energy

EXPECTED BACKGROUNDS

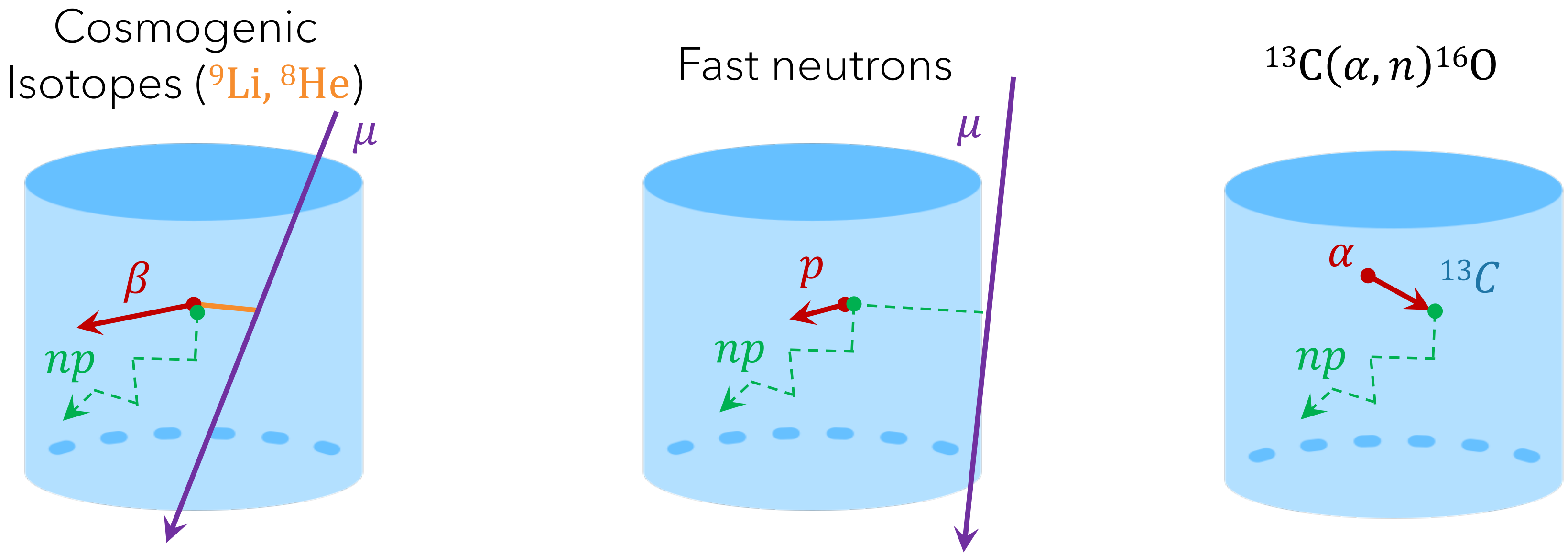
Backgrounds can be divided into two categories:

1. Correlated background: pair of events induced by a single physics process:
    - Geoneutrinos are electron antineutrinos and interact via IBD → irreducible

EXPECTED BACKGROUNDS

Backgrounds can be divided into two categories:

1. **Correlated background:** pair of events induced by a single physics process:
   - Geoneutrinos are electron antineutrinos and interact via IBD → irreducible
   - Cosmogenic isotopes, fast neutrons, $^{13}\mathrm{C}(\alpha, n)^{16}\mathrm{O}$ mimic the prompt-delayed pattern

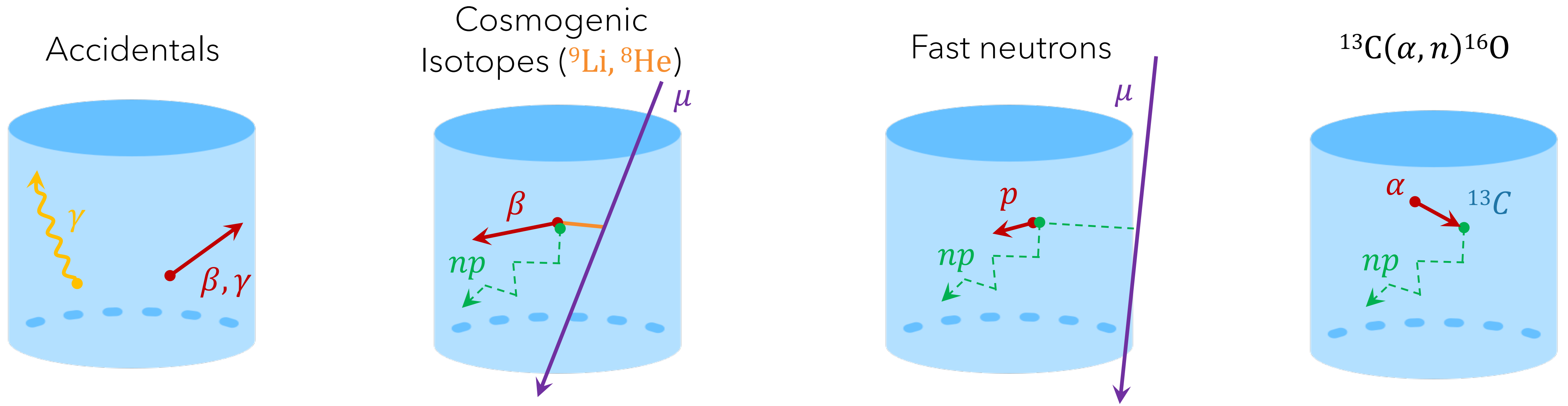


EXPECTED BACKGROUNDS

Backgrounds can be divided into two categories:

1. **Correlated background:** pair of events induced by a single physics process:
   - Geoneutrinos are electron antineutrinos and interact via IBD → irreducible
   - Cosmogenic isotopes, fast neutrons, $^{13}\mathrm{C}(\alpha,n)^{16}\mathrm{O}$ mimic the prompt-delayed pattern

2. **Uncorrelated background:** accidental coincidences
   - Two independent signals (mainly from radioactive contamination) mimic the IBD pattern



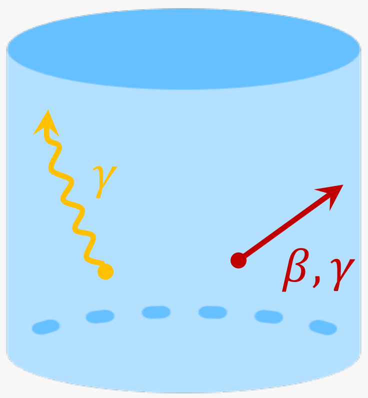
EXPECTED BACKGROUNDS

Backgrounds can be divided into two categories:

1. **Correlated background:** pair of events induced by a single physics process:
   - Geoneutrinos are electron antineutrinos and interact via IBD → irreducible
   - Cosmogenic isotopes, fast neutrons, $^{13}\mathrm{C}(\alpha, n)^{16}\mathrm{O}$ mimic the prompt-delayed pattern

2. **Uncorrelated background:** accidental coincidences
   - Two independent signals (mainly from radioactive contamination) mimic the IBD pattern

Accidentals



Before any selection, rate of single radioactive events is much higher than expected signal rate

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

DATA DESCRIPTION AND BENCHMARK SELECTION

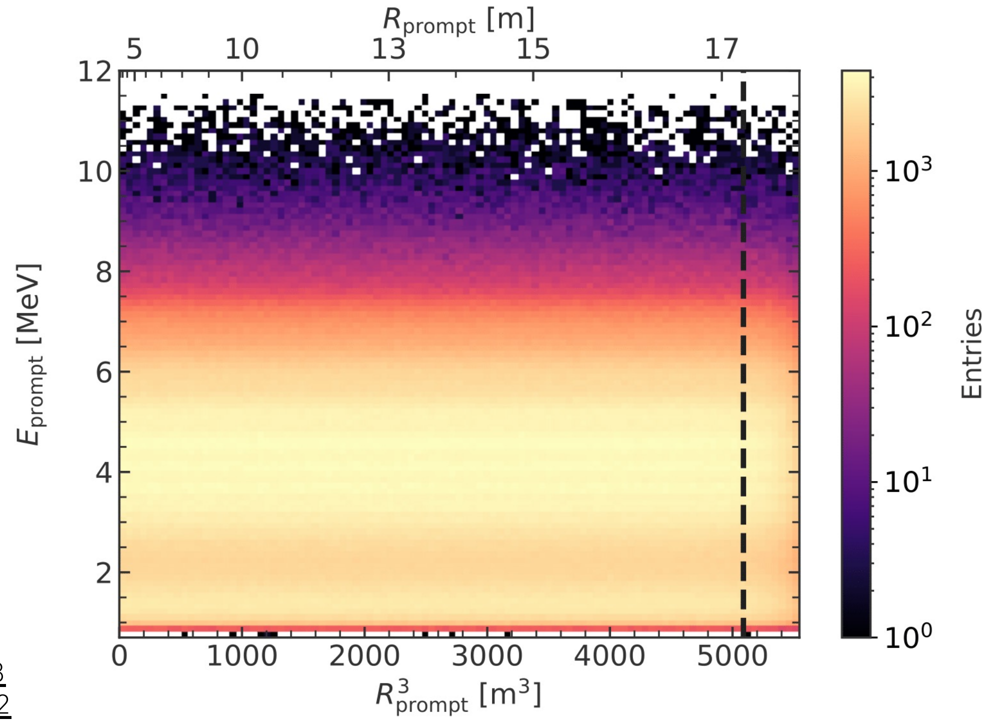
DATA DESCRIPTION

- Highly imbalanced data: 57.4 **IBDs** /day [1] vs >10^5 **accidentals** /day [2] (pairs)

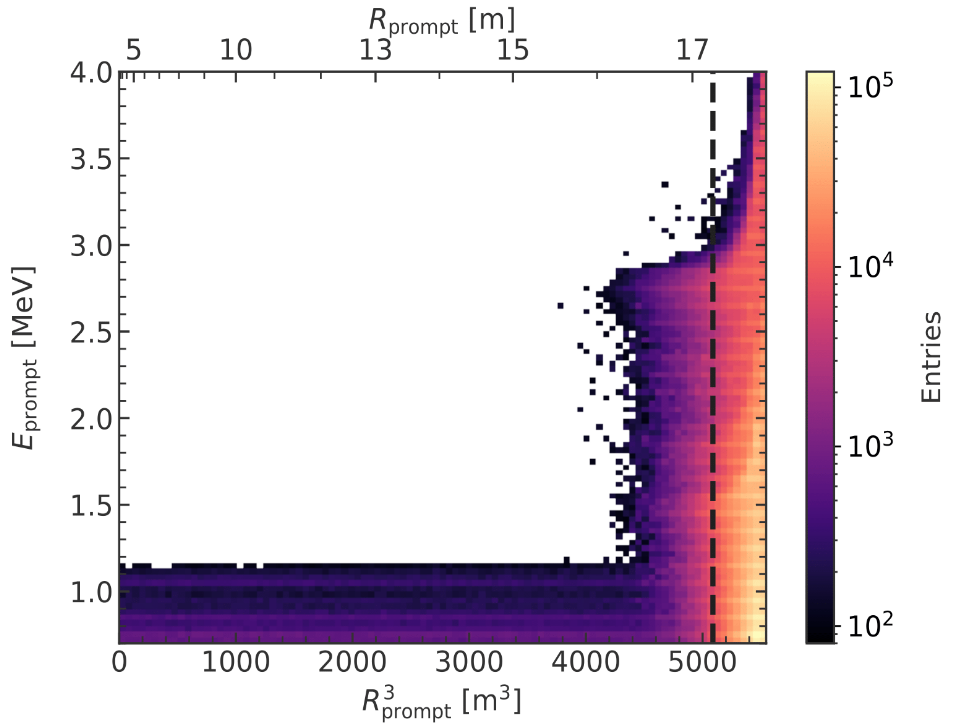
- For training and hyperparameter optimization, **balanced dataset is used** to prevent biases:
  - 15M of IBD pairs
  - 15M of accidental coincidences

20M for training, 5M for validation, 5M for testing

IBDs (signal): *Uniformly distributed in the detector*

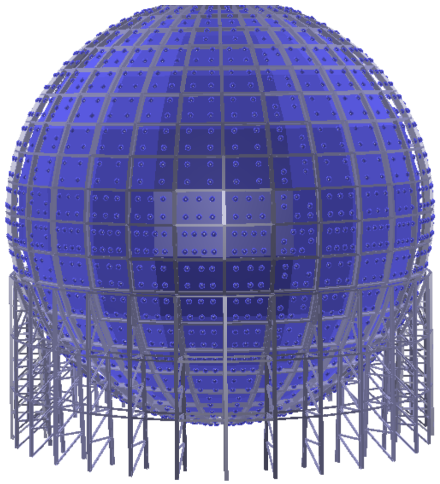


Accidentals (background): *Mostly at the detector edge*



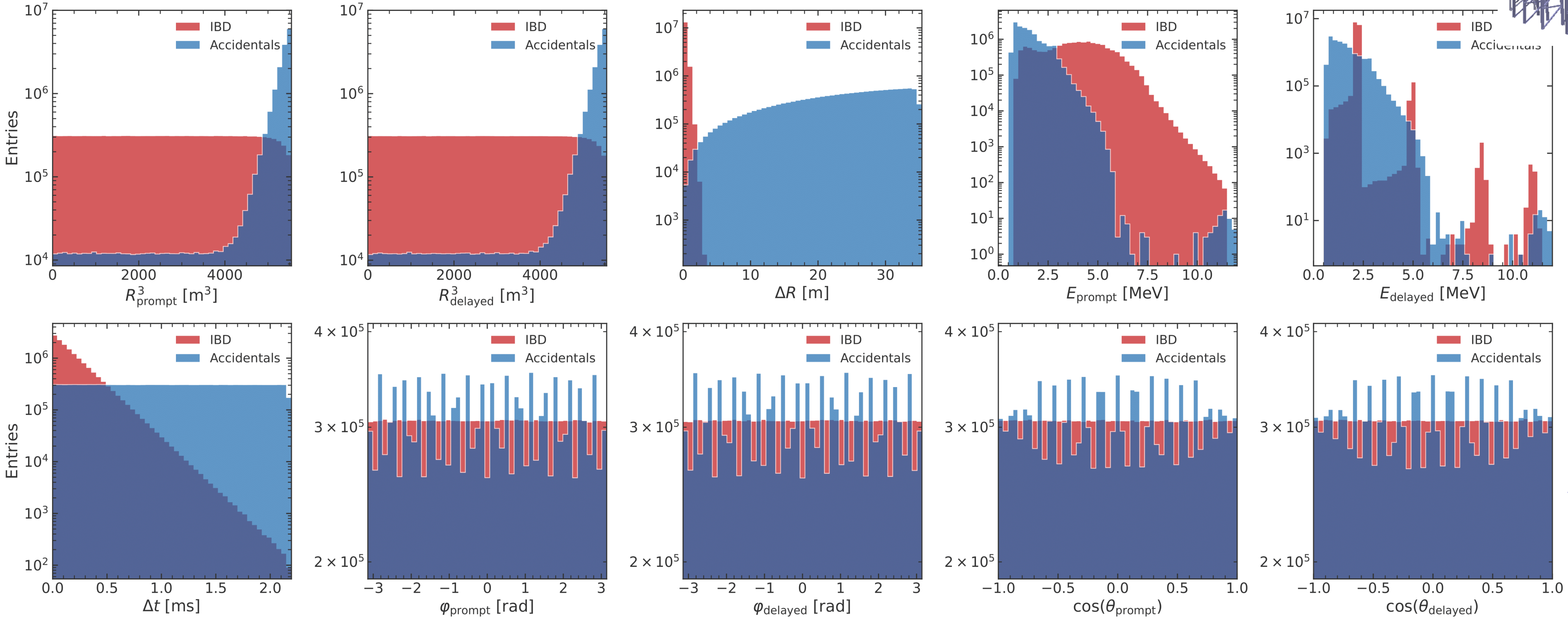
[1] arXiv:2405.18008
[2] JHEP11(2021)102

DATA DESCRIPTION: FEATURE DISTRIBUTIONS



10 energy and vertex related features:

$$R^3{}_p, R^3{}_d, \Delta R, \Delta t, Ep, E_d$$
$$\cos\theta_p, \cos\theta_d, \varphi_p, \varphi_d$$



Higher radioactivity at the edges from the stainless steel grid

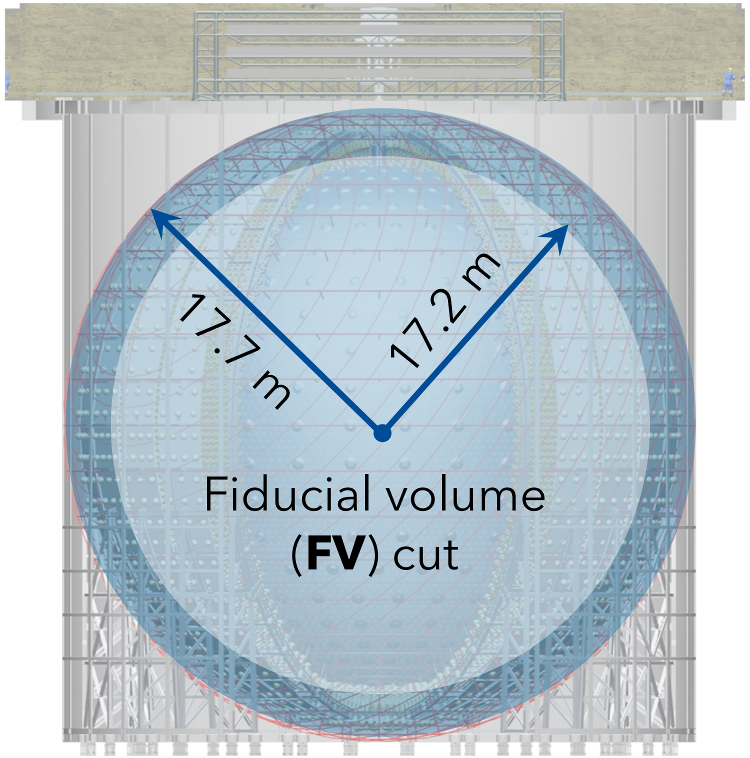
*The dataset was produced independently of the JUNO official software

BENCHMARK SELECTION: CUTS

✶ Benchmark for a machine learning approach → cut-based selection strategy from [1]

Independent cuts

		Efficiency
FV cut	$R < 17.2$ m	**91.7%**
Energy cut	$E_d \in (1.9, 2.5) \cup (4.4, 5.5)$ MeV $E_p \in (0.7, 12)$ MeV	**98.7%**
Vertex cut	$\Delta R < 1.5$ m	**99.0%**
Time cut	$\Delta t < 1$ ms	**99.4%**



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

[1] arXiv:2405.18008

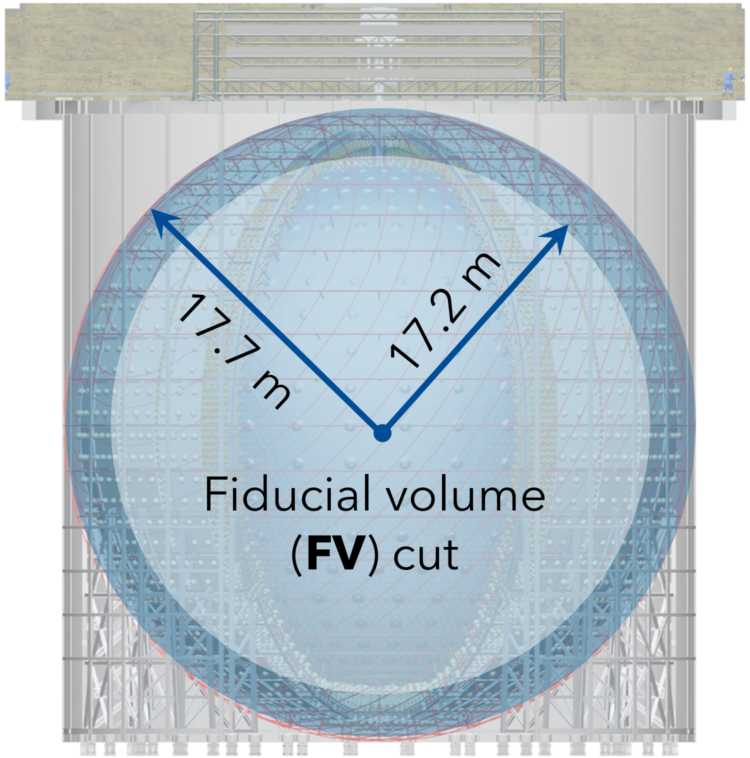
BENCHMARK SELECTION: CUTS

- Benchmark for a machine learning approach → cut-based selection strategy from [1]

- Room for improvement in standard selection
  - Can we **remove the FV cut and retain border events** keeping the same purity?

Independent cuts

		Efficiency
FV cut	$R < 17.2\ \mathrm{m}$	**91.7%**
Energy cut	$E_d \in (1.9, 2.5) \cup (4.4, 5.5)$ MeV $E_p \in (0.7, 12)$ MeV	**98.7%**
Vertex cut	$\Delta R < 1.5\ \mathrm{m}$	**99.0%**
Time cut	$\Delta t < 1\ \mathrm{ms}$	**99.4%**



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

[1] arXiv:2405.18008

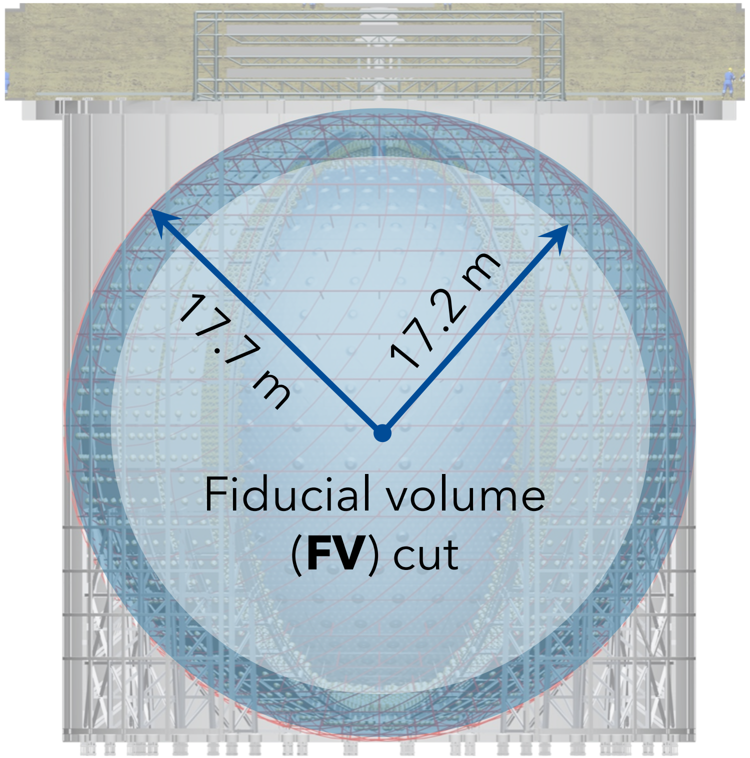
BENCHMARK SELECTION: CUTS

- Benchmark for a machine learning approach → cut-based selection strategy from [1]
- Room for improvement in standard selection
  - Can we **remove the FV cut and retain border events** keeping the same purity?
  - Can we **increase the efficiency within FV**?

Efficiency

Independent cuts	Cut	Efficiency
FV cut	$R < 17.2$ m	**91.7%**
Energy cut	$E_d \in (1.9, 2.5) \cup (4.4, 5.5)$ MeV $E_p \in (0.7, 12)$ MeV	**98.7%**
Vertex cut	$\Delta R < 1.5$ m	**99.0%**
Time cut	$\Delta t < 1$ ms	**99.4%**

Combined 97.1%



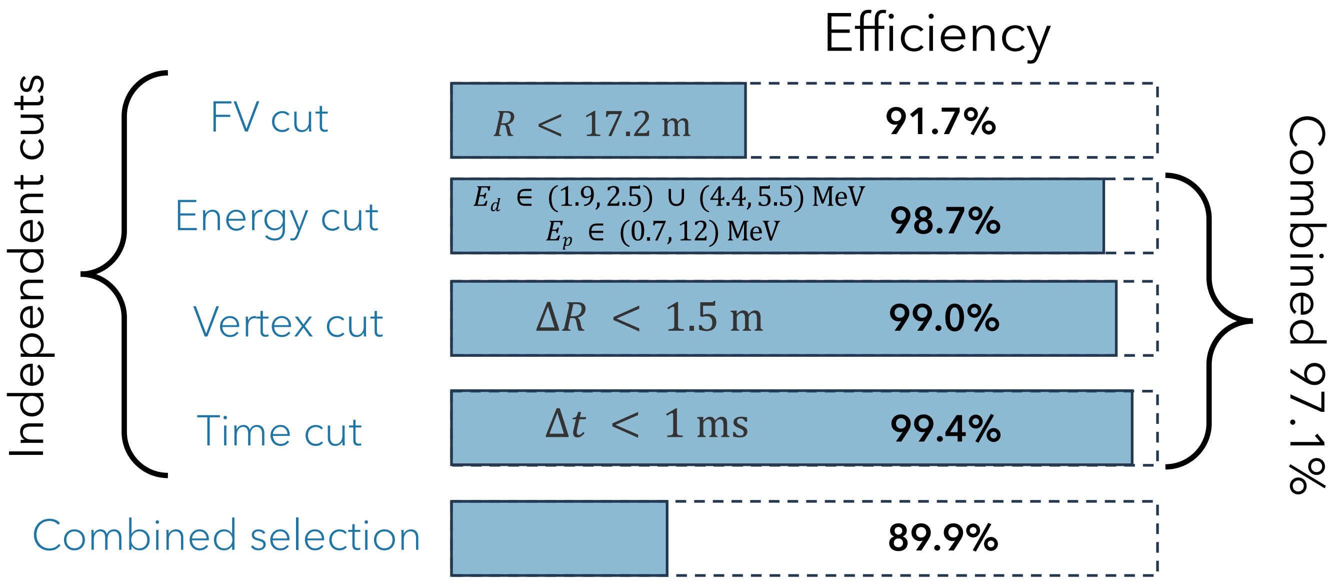
*The cut-based selection does not necessarily reproduce JUNO official selection

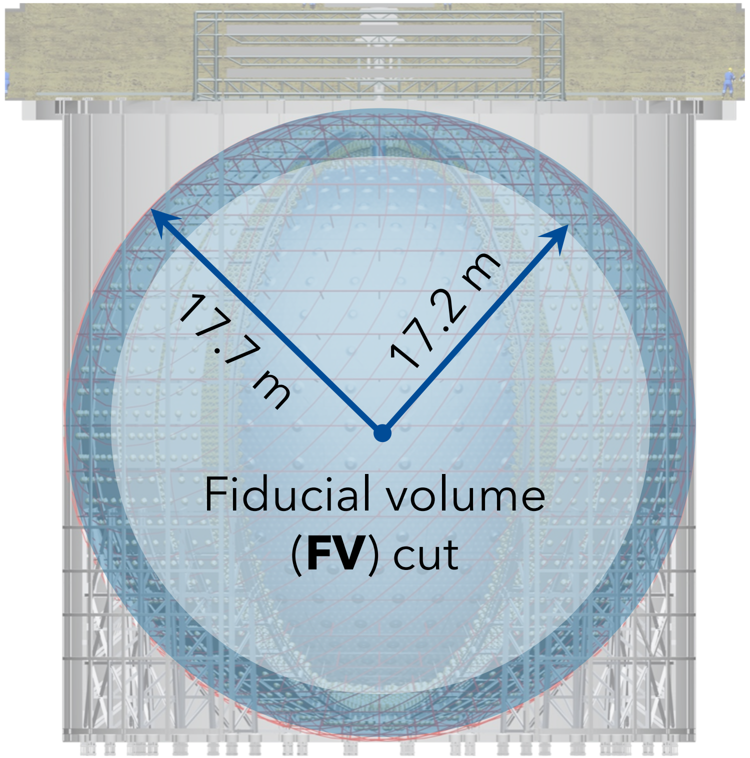
[1] arXiv:2405.18008

BENCHMARK SELECTION: CUTS

- Benchmark for a machine learning approach → cut-based selection strategy from [1]

- Room for improvement in standard selection
  - Can we **remove the FV cut and retain border events** keeping the same purity?
  - Can we **increase the efficiency within FV**?
  - Can we **improve the overall efficiency**?





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

[1] arXiv:2405.18008

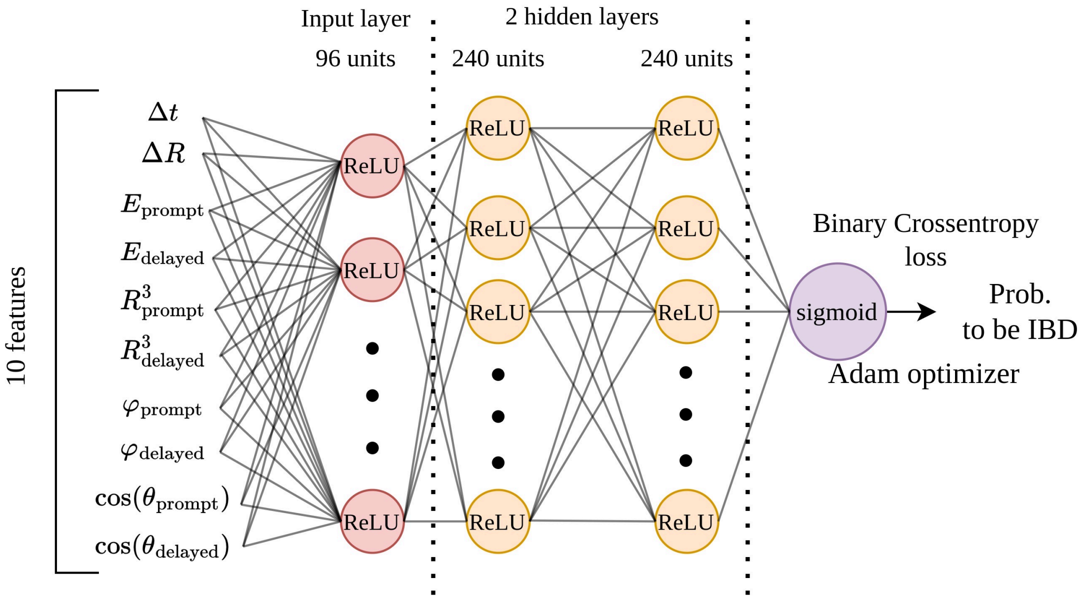
MACHINE LEARNING APPROACH

NEURAL NETWORK ARCHITECTURE

- ✶ Supervised classification problem
- ✶ Goal: **separate** between IBDs and accidentals
- ✶ Comprehensive **hyperparameter optimization** to find optimal number of layers, number of units in a layer, learning rate, etc., …

- ✶ 10 input features:
  - Time and vertex distance: $\Delta t, \Delta R$
  - Prompt and delayed energies: E_p, E_d
  - Distance from detector center: $R^3{}_p, R^3{}_d$
  - Angular variables: $\varphi_p, \varphi_d, \cos\theta_p, \cos\theta_d$

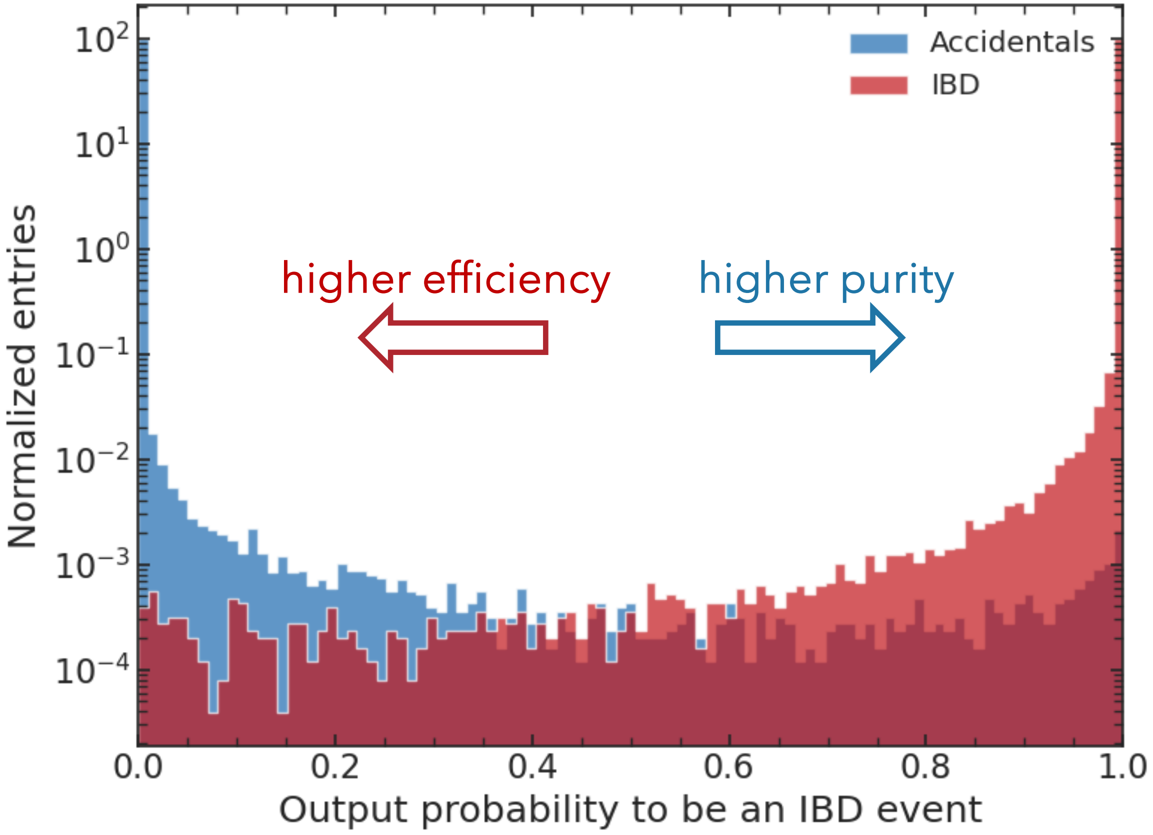
Optimized architecture: Fully Connected Neural Network



*The dataset was produced independently of the JUNO official software

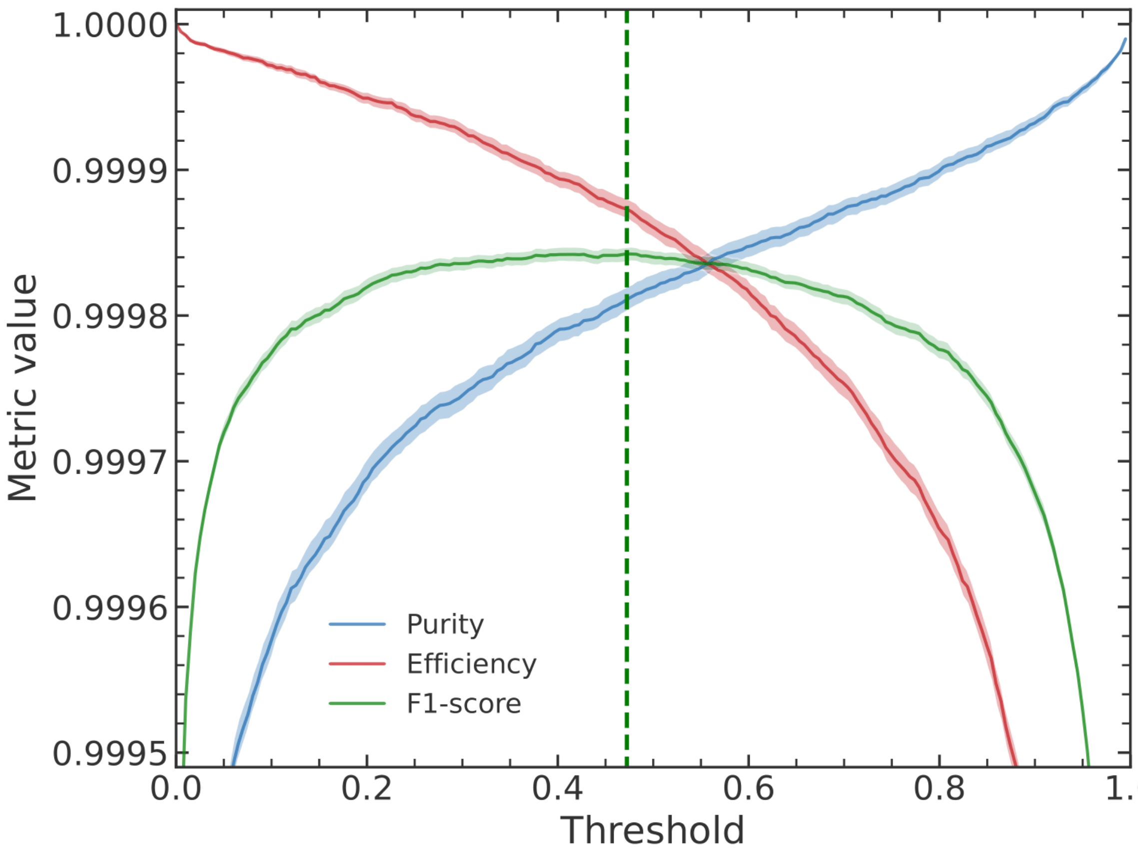
NEURAL NETWORK: A TUNABLE CLASSIFIER

- Output: **confidence score to be an IBD event**, from 0 to 1



NEURAL NETWORK: A TUNABLE CLASSIFIER

- ✶ Output: **confidence score 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:
    - same model
    - different thresholds
    - optimize the desired metric (efficiency or purity)



Balance purity and efficiency by maximizing the *harmonic mean* of efficiency and purity: **F1-score**

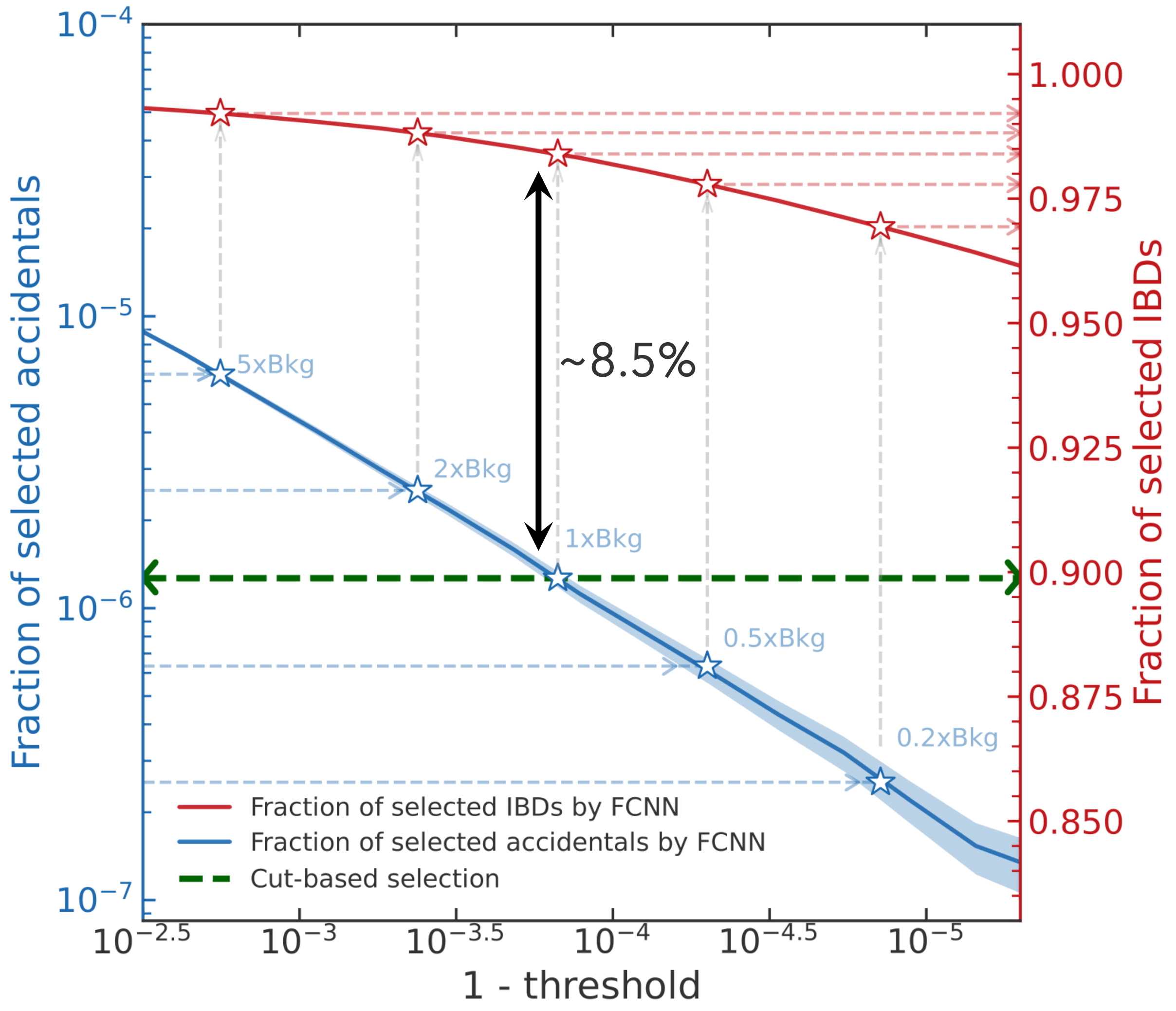
$$\text{F1-score} = 2 \cdot \frac{\text{purity} \cdot \text{efficiency}}{\text{purity} + \text{efficiency}}$$

NEURAL NETWORK: PERFORMANCE

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

- ✶ Improvement of ~8.5% points in efficiency for the same background level as for the cuts*

Unbalanced dataset: 150M accidentals, 2.5M IBDs



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

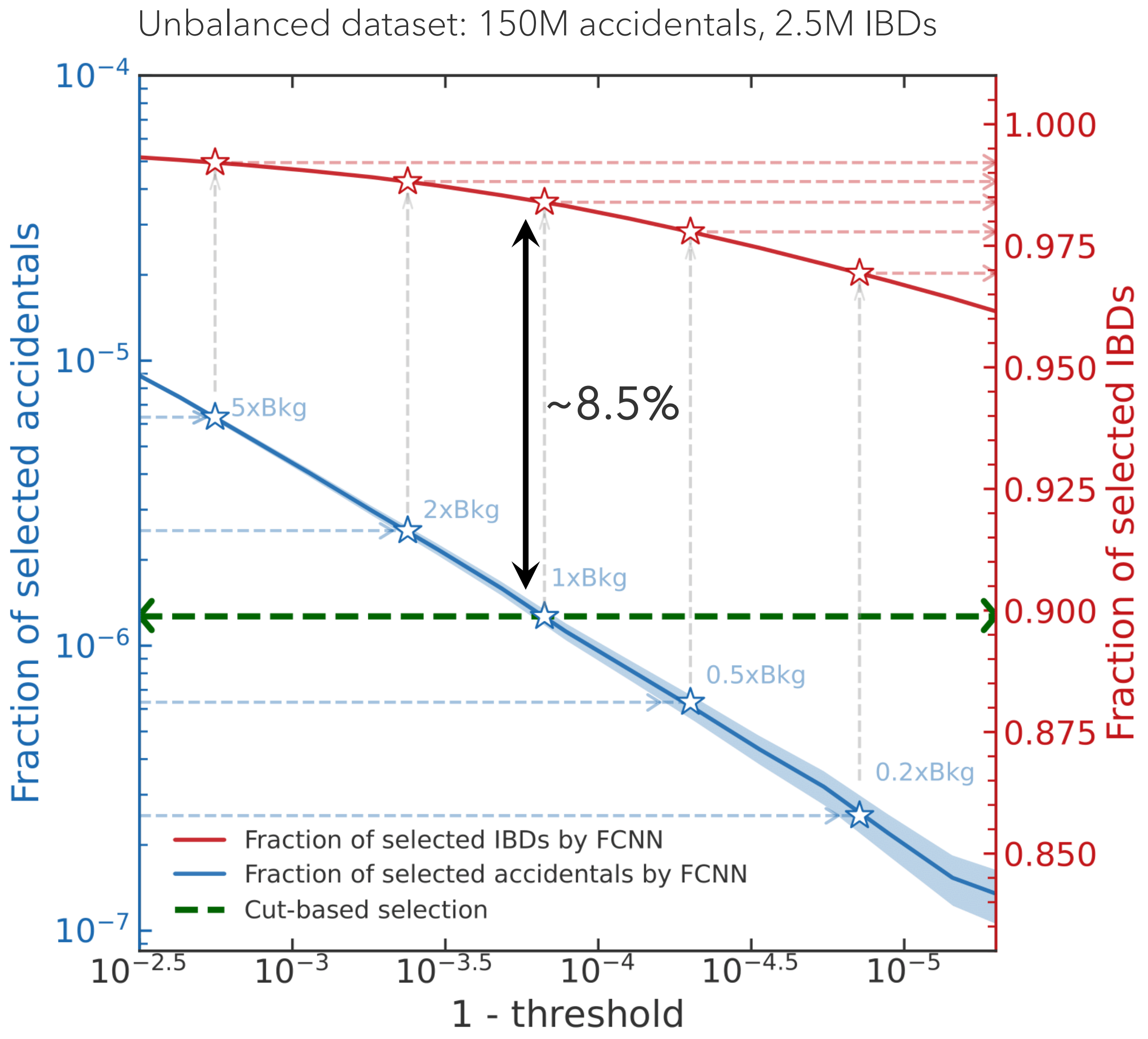
NEURAL NETWORK: PERFORMANCE

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

- Improvement of ~8.5% points in efficiency for the same background level as for the cuts*

- Background can be further decreased, keeping higher efficiency

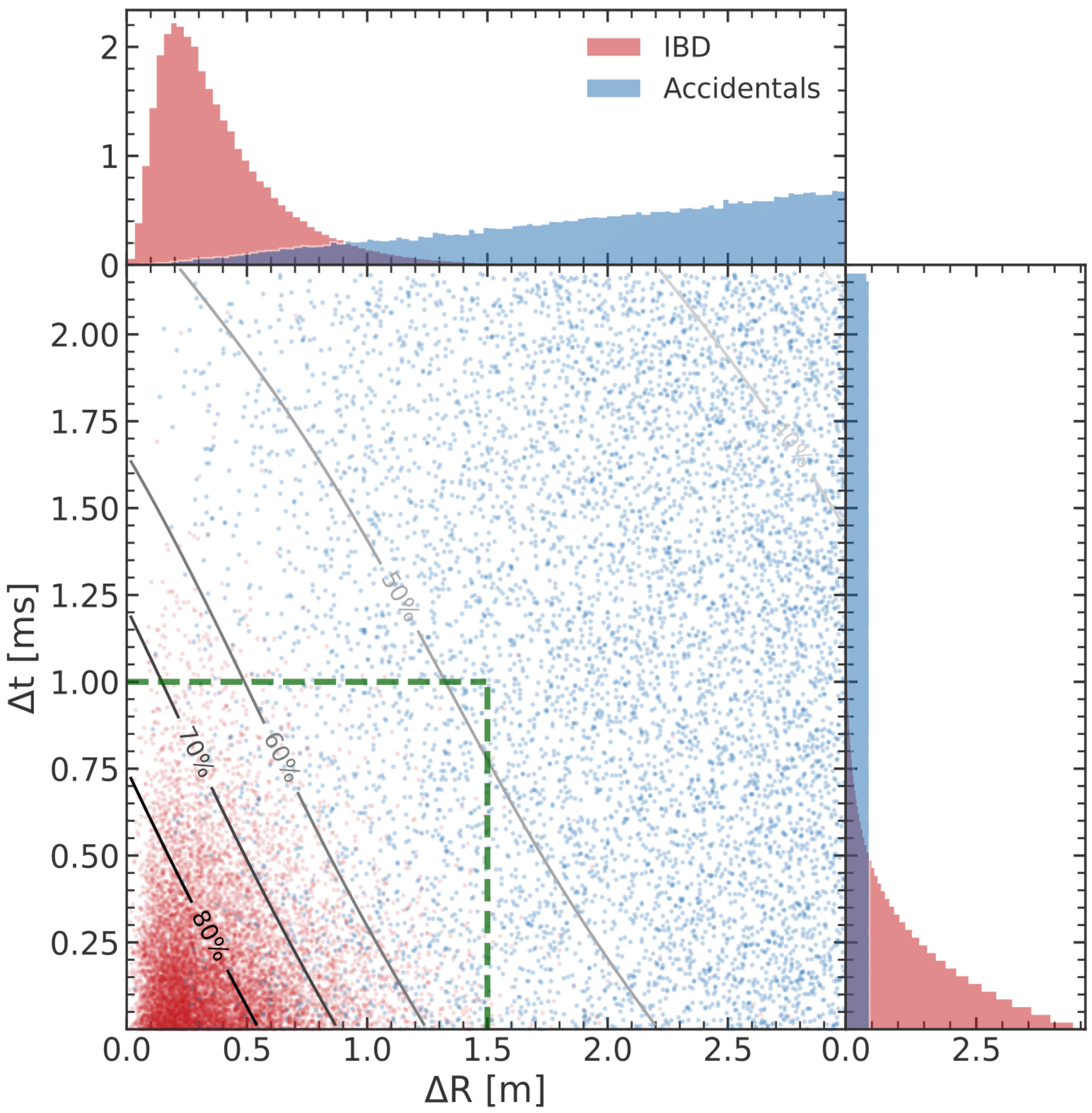
- Efficiency can be further increased, having higher residual background



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

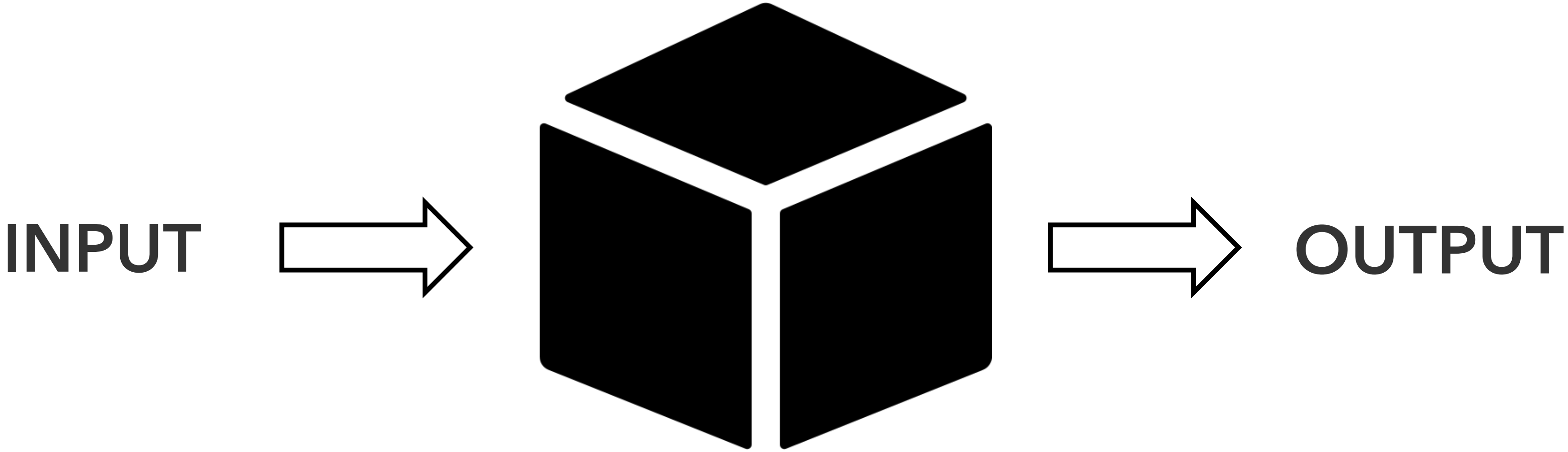
OPTIMIZATION OF IBD SELECTION CUTS

- ✶ Standard ΔR and Δt cuts are box-like
- ✶ ML model has learnt smoother decision boundary
  - Increased efficiency
  - Close-to-linear relation
  - Can be used to optimize selection criteria



INTERPRETABILITY

WHY INTERPRETABILITY?



- ✶ Hard to understand how and why a neural network made a decision → *Black box*

WHY INTERPRETABILITY?

- ✶ Hard to understand how and why a neural network made a decision → ~~*Black box*~~
- ✶ Explainable AI: methods that allow users to comprehend results created by AI algorithms

WHY INTERPRETABILITY?

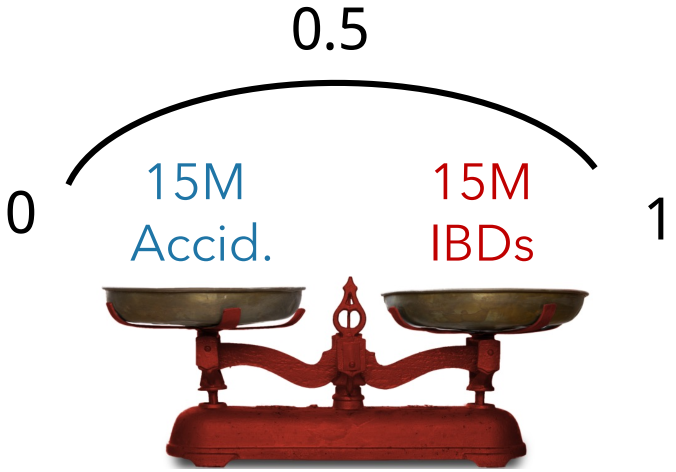
- ✶ Hard to understand how and why a neural network made a decision → ~~*Black box*~~
- ✶ Explainable AI: methods that allow users to comprehend results created by AI algorithms
- ✶ Goals:
    - **Ensure trust** in the model and its **transparency**
    - Identify features driving decision about classification
    - **Optimize and fine-tune** the cut-based selection strategy

INTERPRETABILITY AND SHAP VALUES

- SHAP (**SH**apley **A**dditive ex**P**lanations)
  - Calculate the contribution of each feature
  - Can be positive or negative
  - Show the **impact on predictions with respect to the average**

- Positive SHAP values → IBD class

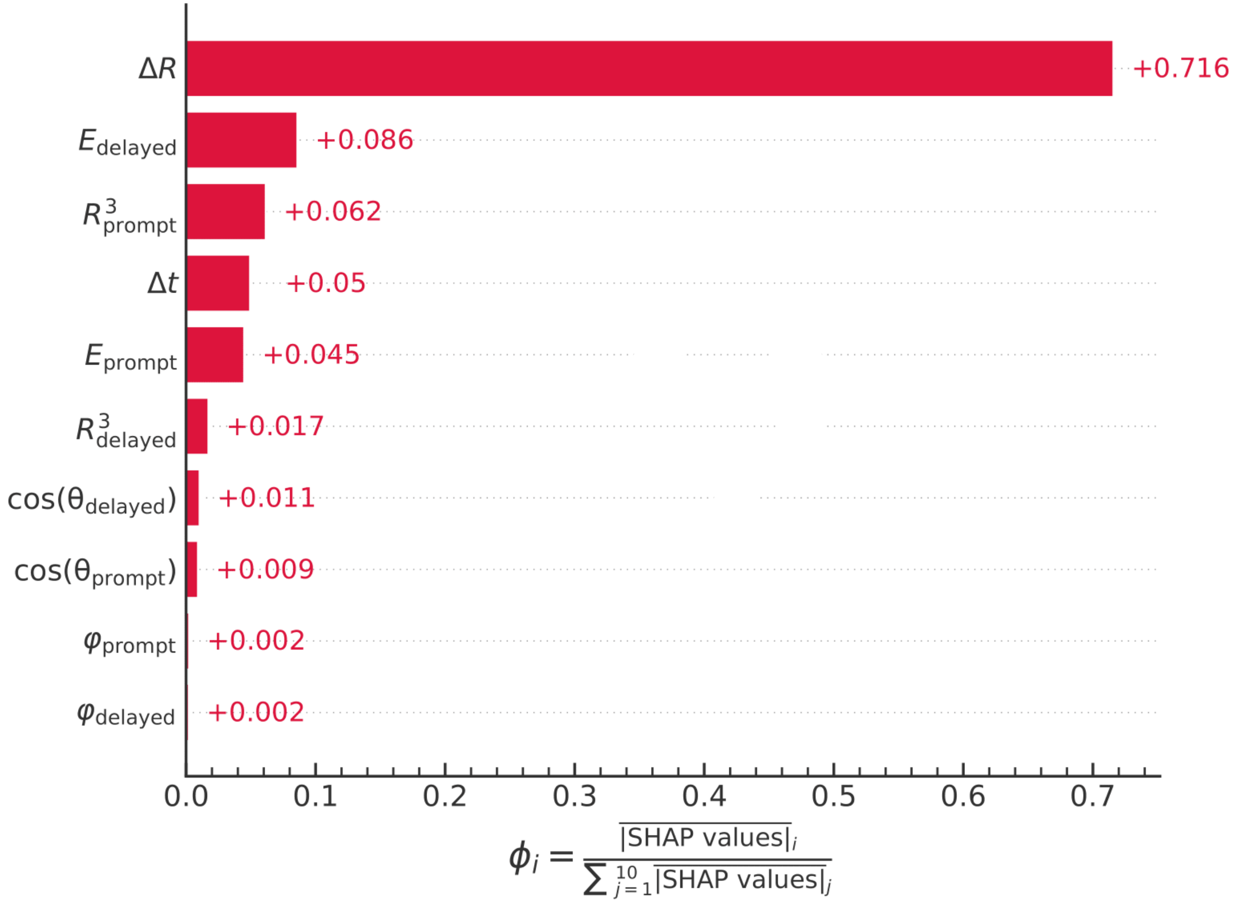
- Negative SHAP values → accidentals class



GLOBAL EXPLANATIONS

What features are the most important for the model's predictions on average?

- Global explanations: **summarized impact** of a specific feature
- Most important feature is distance between prompt and delayed candidates' vertices ΔR
- Energy of delayed candidate has **the next stronger** discriminative power
- $R^3{}_{\text{prompt}}$, $R^3{}_{\text{delayed}}$ provide additional information, especially at the edges
- Δt in top 4 features for importance



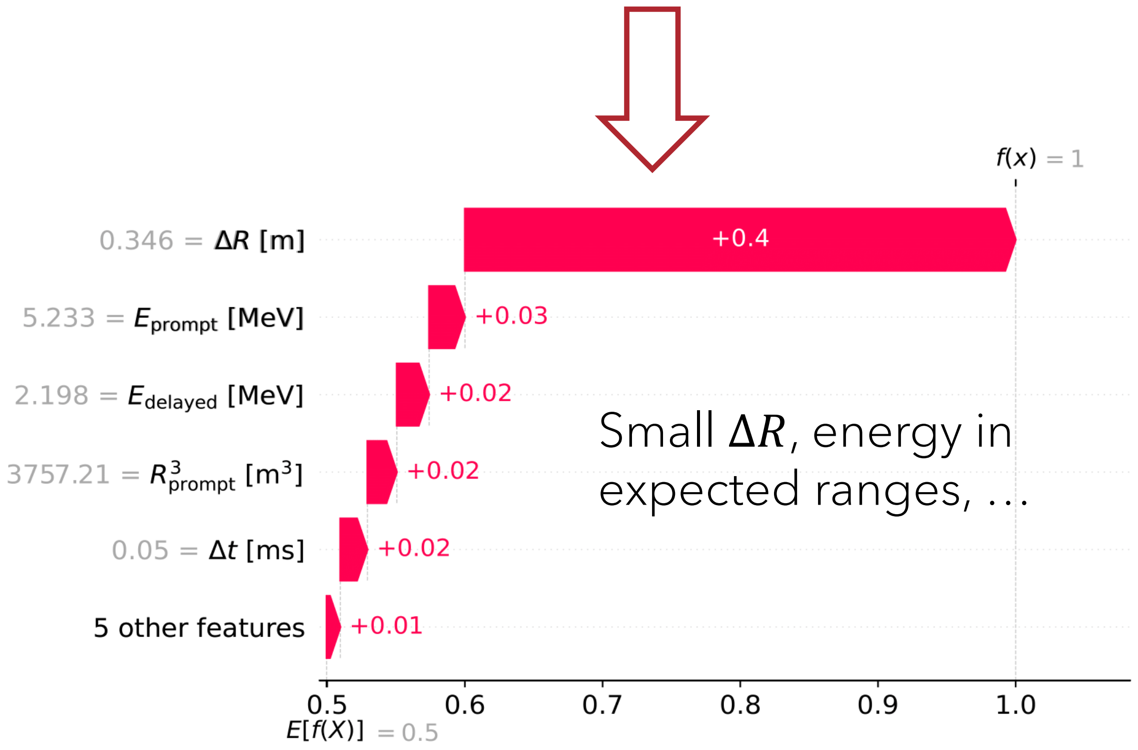
LOCAL EXPLANATIONS – GOOD CLASSIFICATION

Why did the model make this prediction for this specific data point?

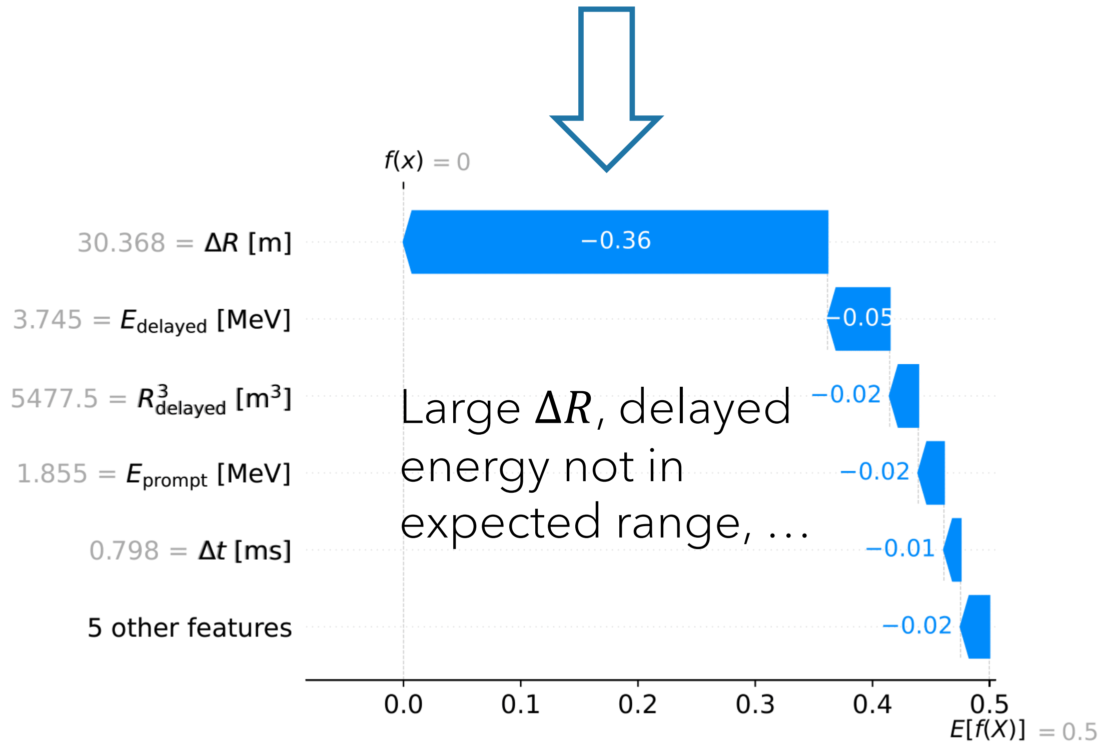
- ✶ Local explanations focus on an individual event and provide features' importances for a specific instance

- ✶ **Example of good classification**: typical correctly classified events

True positives (IBDs)



True negatives (accidentals)

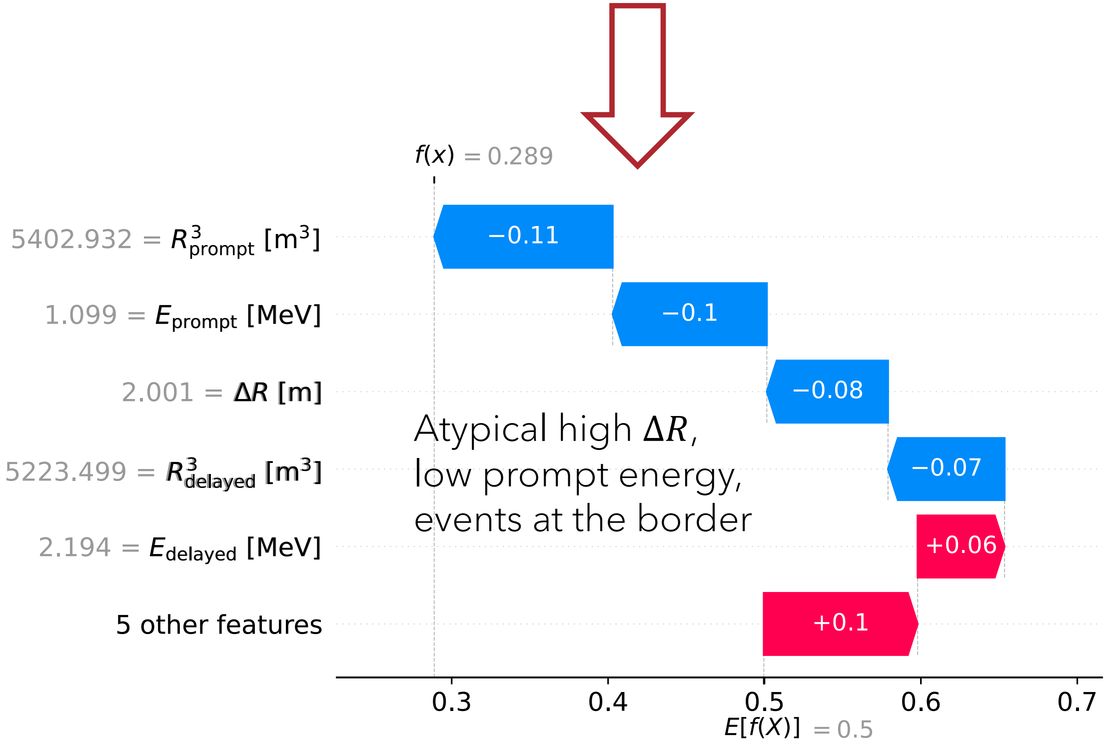


LOCAL EXPLANATIONS – MIS-CLASSIFICATION

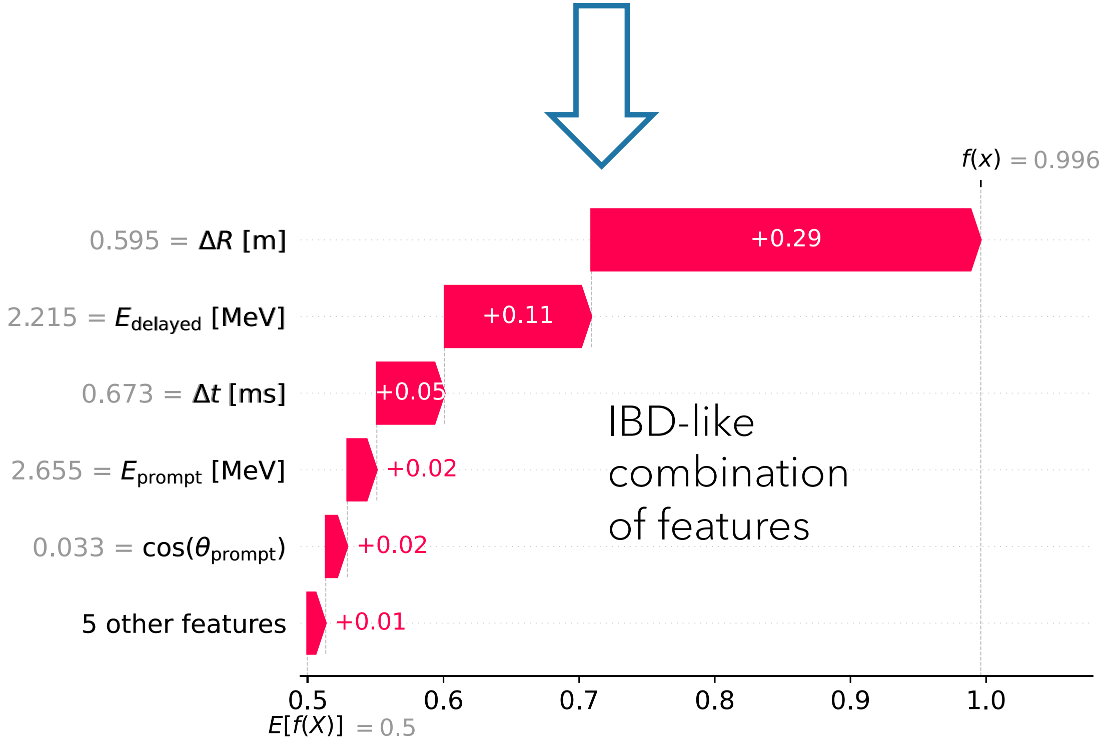
- Local explanations focus on an individual event and provide features' importances for a specific instance

- **Example of mis-classification**: wrongly classified events

IBDs classified as accidentals



Accidentals classified as IBDs

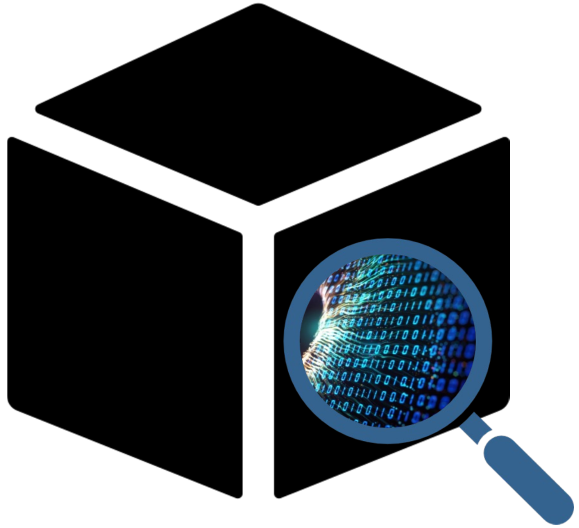


CONCLUSIONS

- ✶ The main task of a **selection algorithm** is to distinguish between two classes: **reactor antineutrino events and accidental coincidences** (mainly from radioactivity)

- ✶ Introduced a fully-connected **neural network** as IBD-accidentals classifier
  - Can be applied as **a proxy to improve the cut-based selection**
  - **Provides higher efficiency** than the cut-based selection within both fiducial volume and the entire volume
  - Provides **higher efficiency** for **higher purity level**

- ✶ Interpretability analysis
  - Ensures trust in the model predictions → ~~*Black box*~~
  - Identify features driving decision about classification
  - **Offers valuable insights** into the model behavior



BACKUP

LS-BASED REACTOR NEUTRINO EXPERIMENTS

- ✶ The medium of choice for most reactor neutrino experiments has been **liquid scintillator (LS)**
- ✶ Organic liquid scintillator detectors:
  - High light yield → energy resolution
  - Large proton abundance → antineutrino target
  - Large volume → compensate small cross section
  - Doping capabilities → improve signal-background discrimination

In this talk

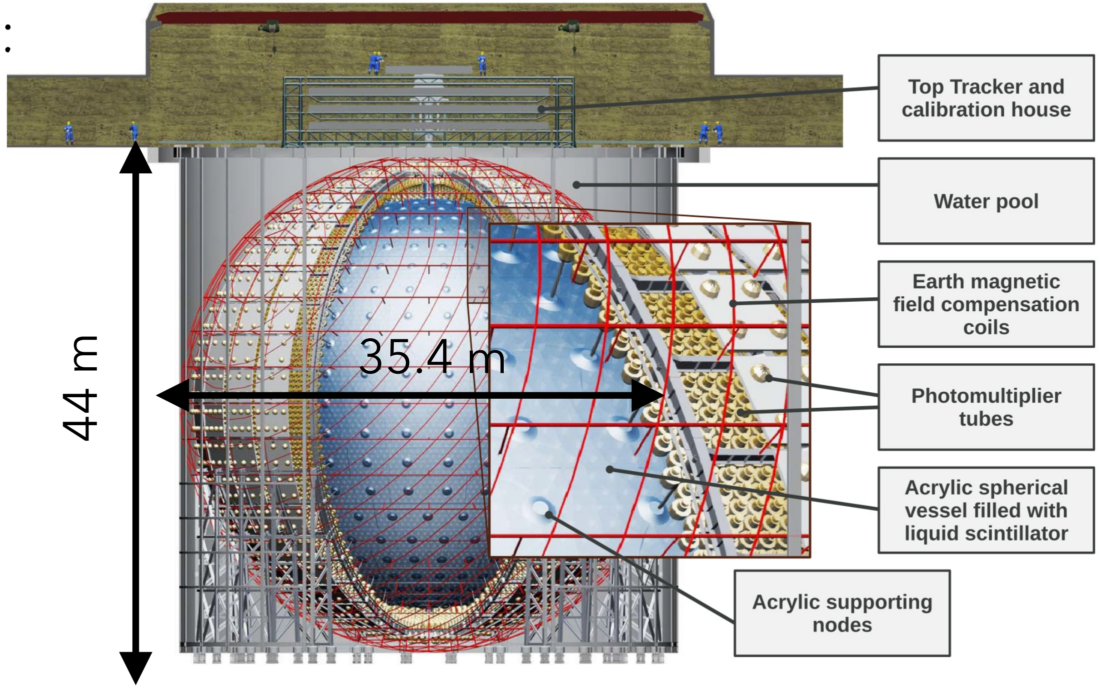
Experiment	Daya Bay	RENO	Double Chooz	KamLAND	JUNO
LS mass	20 ton	16 ton	8 ton	1 kton	20 kton
Energy resolution	8%	8%	8%	6%	3%

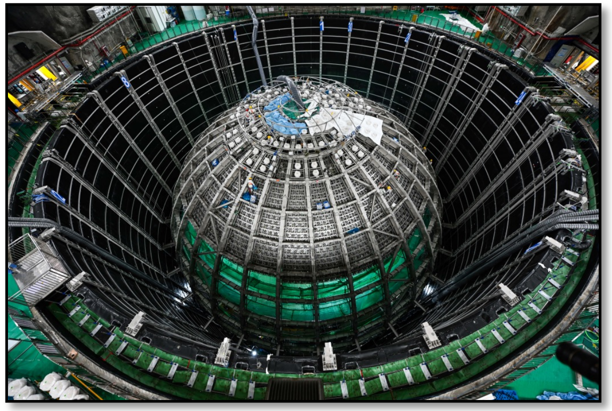
JUNO KEY EXPERIMENTAL FEATURES

PPNP 123 (2022): 103927

Detector design is driven by ambitious physics goals:

- ✶ Large statistics
  - ✓ Huge LS target mass
  - ✓ Powerful nuclear reactors



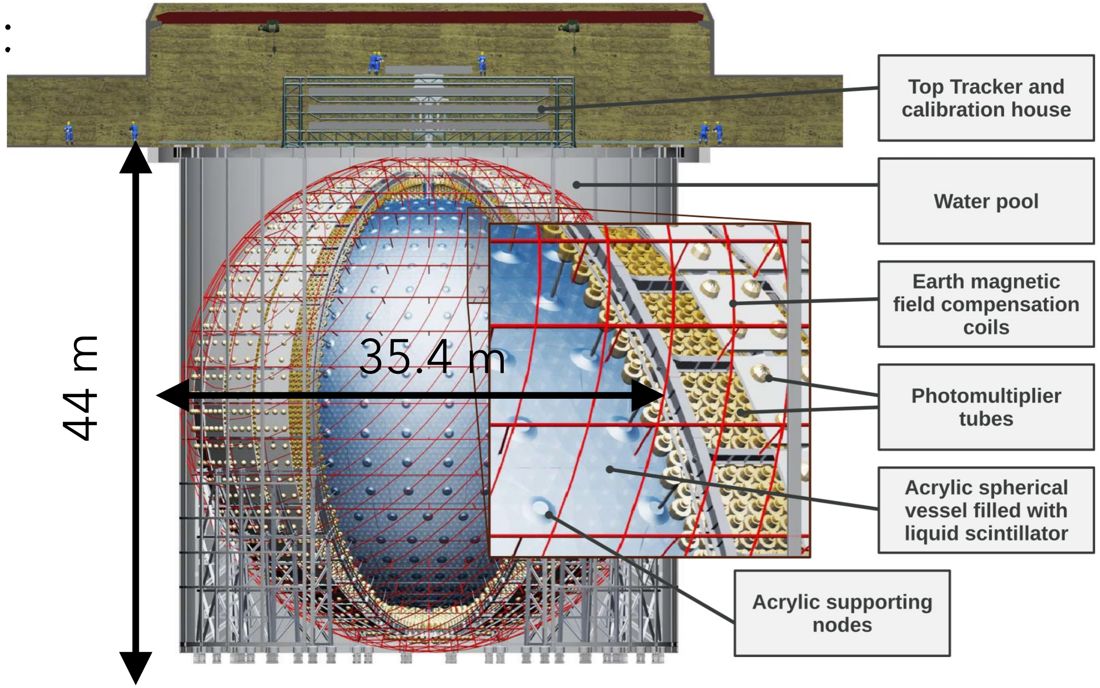


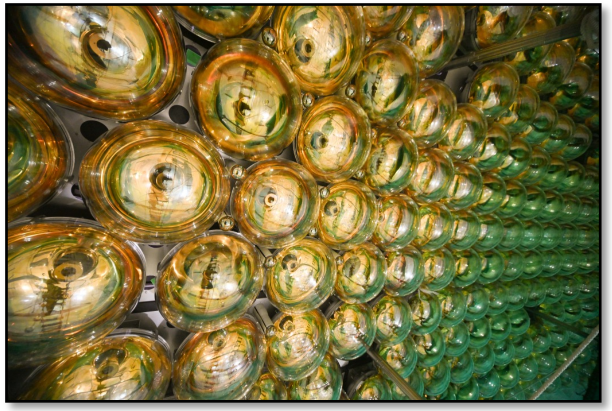
JUNO KEY EXPERIMENTAL FEATURES

PPNP 123 (2022): 103927

Detector design is driven by ambitious physics goals:

- ✶ Large statistics
  - ✓ Huge LS target mass
  - ✓ Powerful nuclear reactors
- ✶ Energy resolution: 2.95% at 1 MeV
  - ✓ High photon yield, highly transparent LS
  - ✓ Total photo-coverage ≈ 78%
- ✶ Control of non-linear energy scale within 1%
  - ✓ Comprehensive calibration program
- ✶ Low background
  - ✓ 650 m underground
  - ✓ LS purification system and material screening
  - ✓ Efficient veto systems



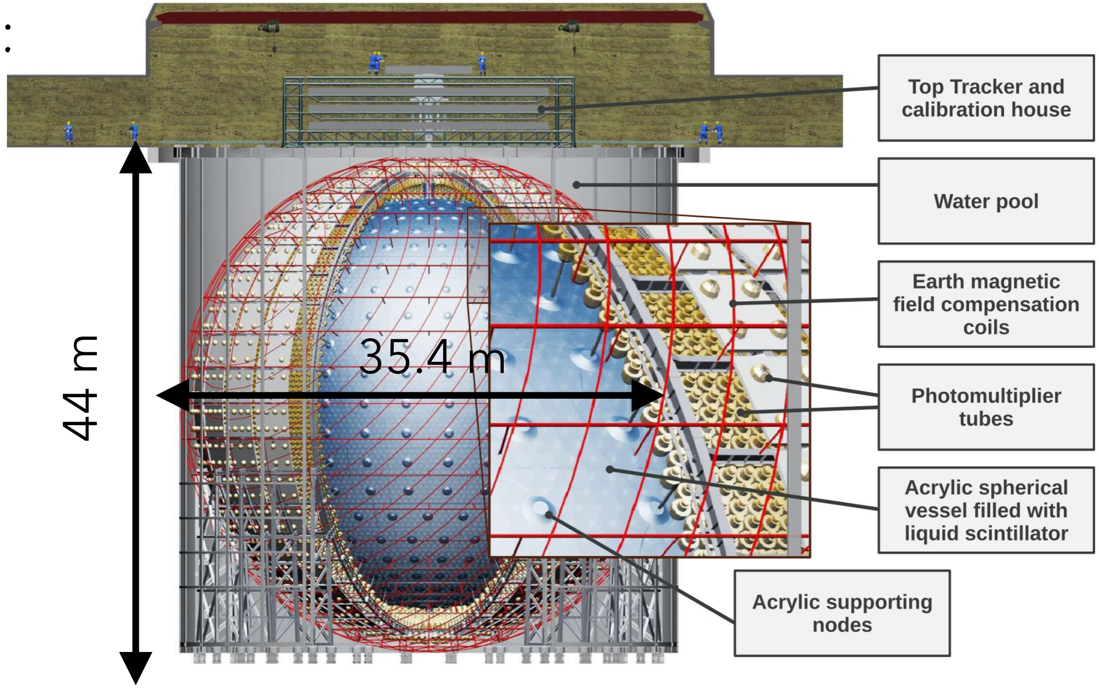


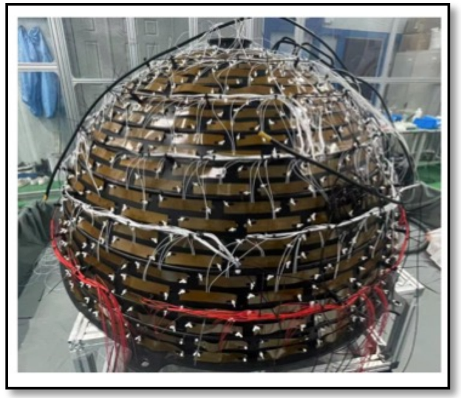
JUNO KEY EXPERIMENTAL FEATURES

PPNP 123 (2022): 103927

Detector design is driven by ambitious physics goals:

- ✶ Large statistics
  - ✓ Huge LS target mass
  - ✓ Powerful nuclear reactors
- ✶ Energy resolution: 2.95% at 1 MeV
  - ✓ High photon yield, highly transparent LS
  - ✓ Total photo-coverage ≈ 78%
- ✶ Control of non-linear energy scale within 1%
  - ✓ Comprehensive calibration program
- ✶ Low background
  - ✓ 650 m underground
  - ✓ LS purification system and material screening
  - ✓ Efficient veto systems
- ✶ Accurate knowledge of reactor spectra
  - ✓ Satellite near detector: **Taishan Antineutrino Observatory (TAO)** at 44 m from Taishan reactor



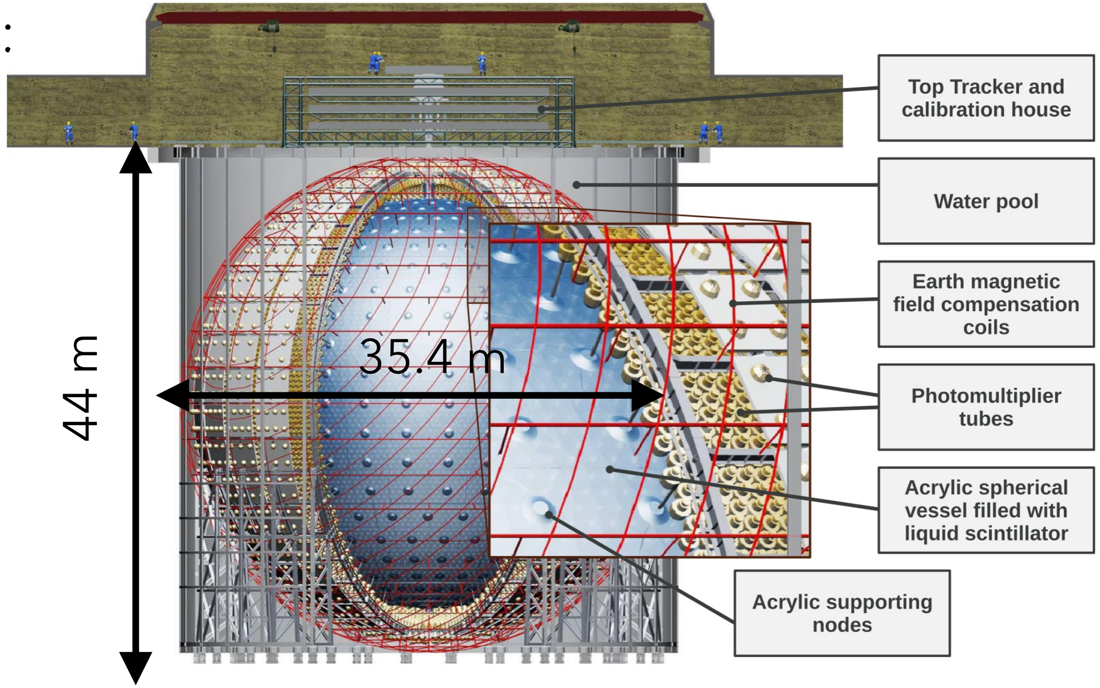


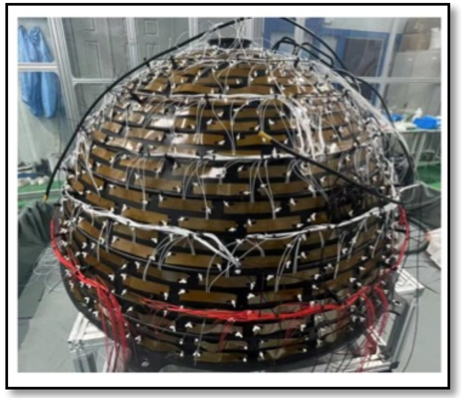
JUNO KEY EXPERIMENTAL FEATURES

PPNP 123 (2022): 103927

Detector design is driven by ambitious physics goals:

- ✶ Large statistics
  - ✓ Huge LS target mass
  - ✓ Powerful nuclear reactors
- ✶ Energy resolution: 2.95% at 1 MeV
  - ✓ High photon yield, highly transparent LS
  - ✓ Total photo-coverage ≈ 78%
- ✶ Control of non-linear energy scale within 1%
  - ✓ Comprehensive calibration program
- ✶ Low background
  - ✓ 650 m underground
  - ✓ LS purification system and material screening
  - ✓ Efficient veto systems
- ✶ Accurate knowledge of reactor spectra
  - ✓ Satellite near detector: **Taishan Antineutrino Observatory (TAO)** at 44 m from Taishan reactor





DETECTOR RESPONSE: WHAT JUNO MEASURES

[Chinese Phys. C 46 123001]

$E_\nu \xrightarrow{\textit{Interaction}} E_{\mathrm{dep}} \xrightarrow{\textit{Light emission}} E_{\mathrm{vis}} \xrightarrow{\textit{Light detection}} E_{\mathrm{rec}}$

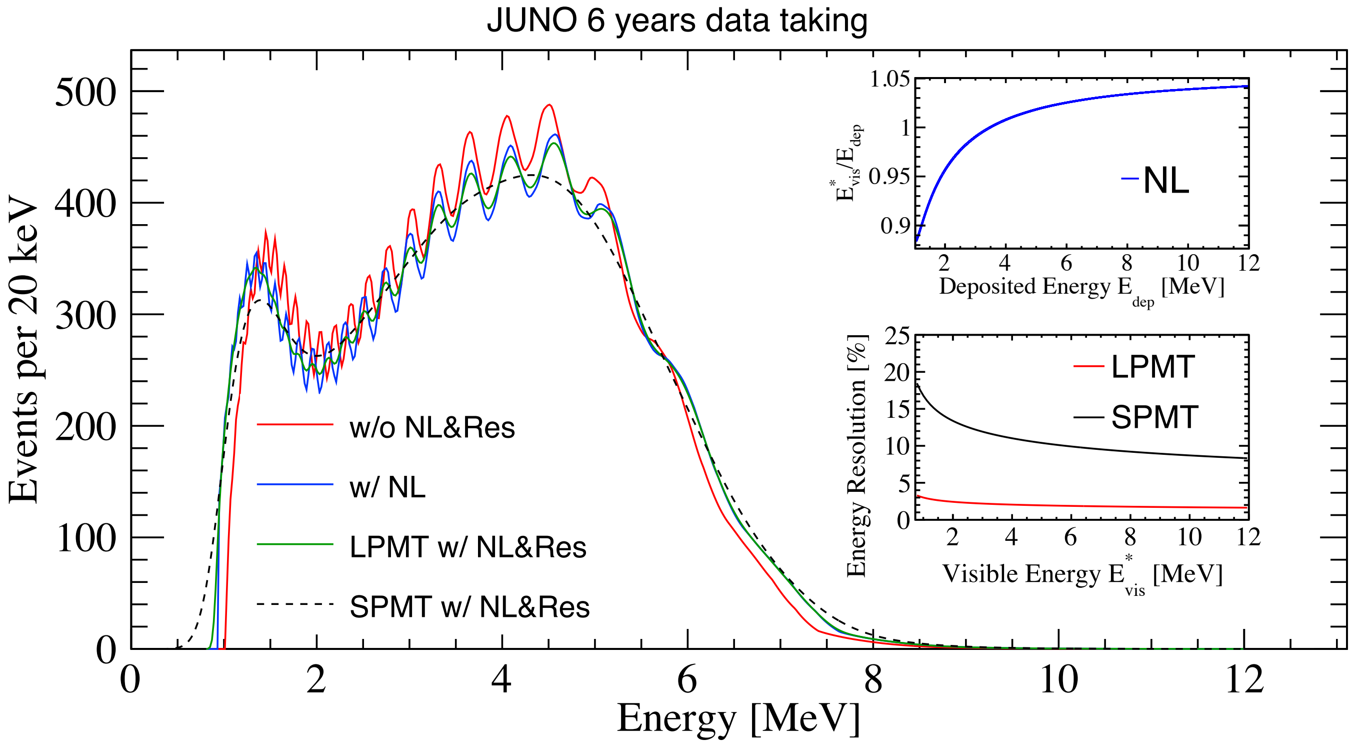
Antineutrino energy	Deposited energy	Visible energy	Reconstructed energy

DETECTOR RESPONSE: WHAT JUNO MEASURES

Chinese Phys. C 46 123001

$E_\nu \xrightarrow{\text{Interaction}} E_{\text{dep}} \xrightarrow{\text{Light emission}} E_{\text{vis}} \xrightarrow{\text{Light detection}} E_{\text{rec}}$

Antineutrino energy — Deposited energy — Visible energy — Reconstructed energy



1. IBD reaction and cross section, e^+ deposited energy

$$E_{\text{dep}} \simeq E_{\bar{\nu}_e} - 0.782\,\text{MeV}$$

2. Liquid scintillator non-linearity (NL), visible energy ∝ detected photoelectrons

$$E_{\text{vis}} = f_{\text{LSNL}}(E_{\text{dep}}) \cdot E_{\text{dep}}$$

3. Energy resolution (Res)

$$\frac{\sigma_{E_{\text{rec}}}}{E_{\text{vis}}} = \sqrt{\left(\frac{a}{\sqrt{E_{\text{vis}}}}\right)^2 + b^2 + \left(\frac{c}{E_{\text{vis}}}\right)^2}$$

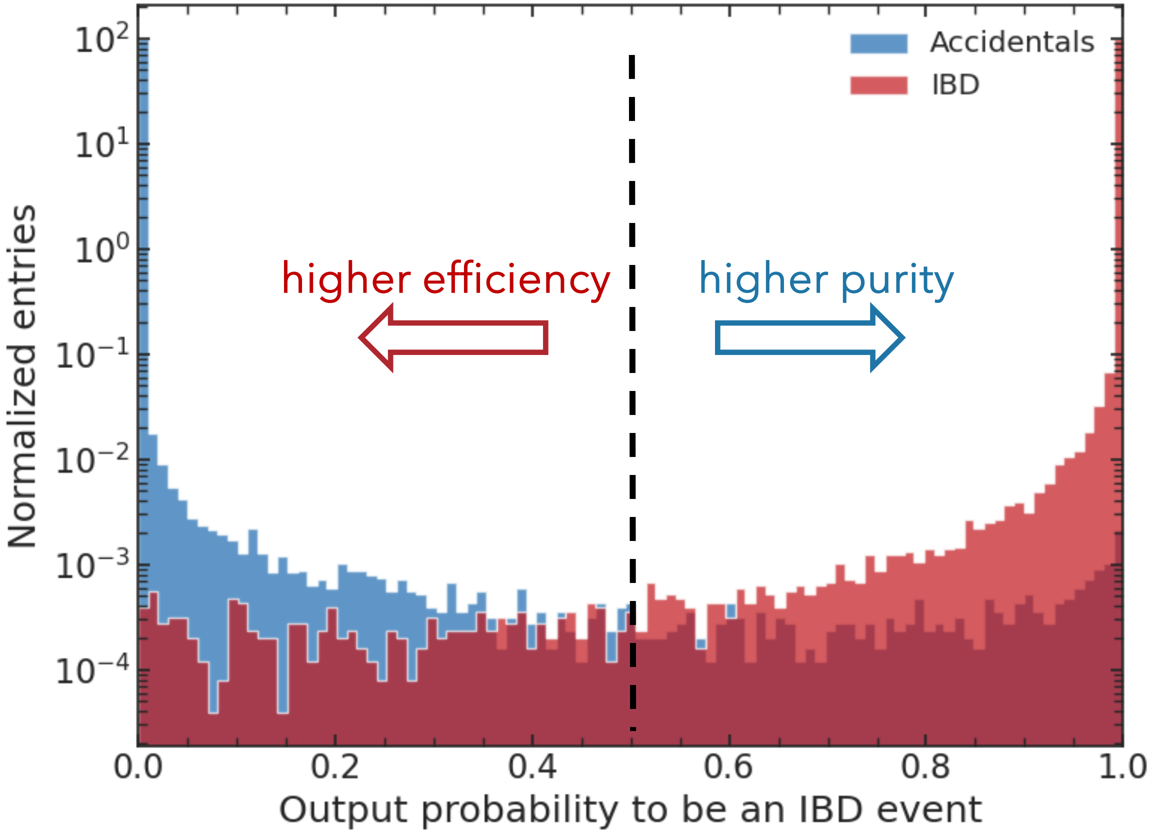
NEURAL NETWORK: A TUNABLE CLASSIFIER

✶ Output: **confidence score 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:

- same model
- different thresholds
- optimize the desired metric (efficiency or purity)



Balance purity and efficiency by maximizing the *harmonic mean* of efficiency and purity: F1-score

$$\text{F1-score} = 2 \cdot \frac{\text{purity} \cdot \text{efficiency}}{\text{purity} + \text{efficiency}}$$

NEURAL NETWORK: HYPERPARAMETER OPTIMIZATION

Hyperparameter	*Search space and **selected hyperparameter***
Units in input layer	[16, 256]: **96**
Units in hidden layers	[16, 256]: **240**
Number of hidden layers	[1, 10]: **2**
Activation	**ReLU**, Leaky ReLU, SiLU, PReLU, Tanh
Optimizer	**Adam**, SGD, RMSprop
Learning rate	$[10^{-5}, 10^{-1}]$: $\mathbf{3.5 \cdot 10^{-4}}$
Scheduler type	Exponential, ReduceOnPlateau, **CosineAnnealing**, None
Layer weights initialization	xavier uniform, xavier normal, **orthogonal**, normal, uniform
Batch normalization	True, **False**
Batch size	[128, 2048]: **1024**

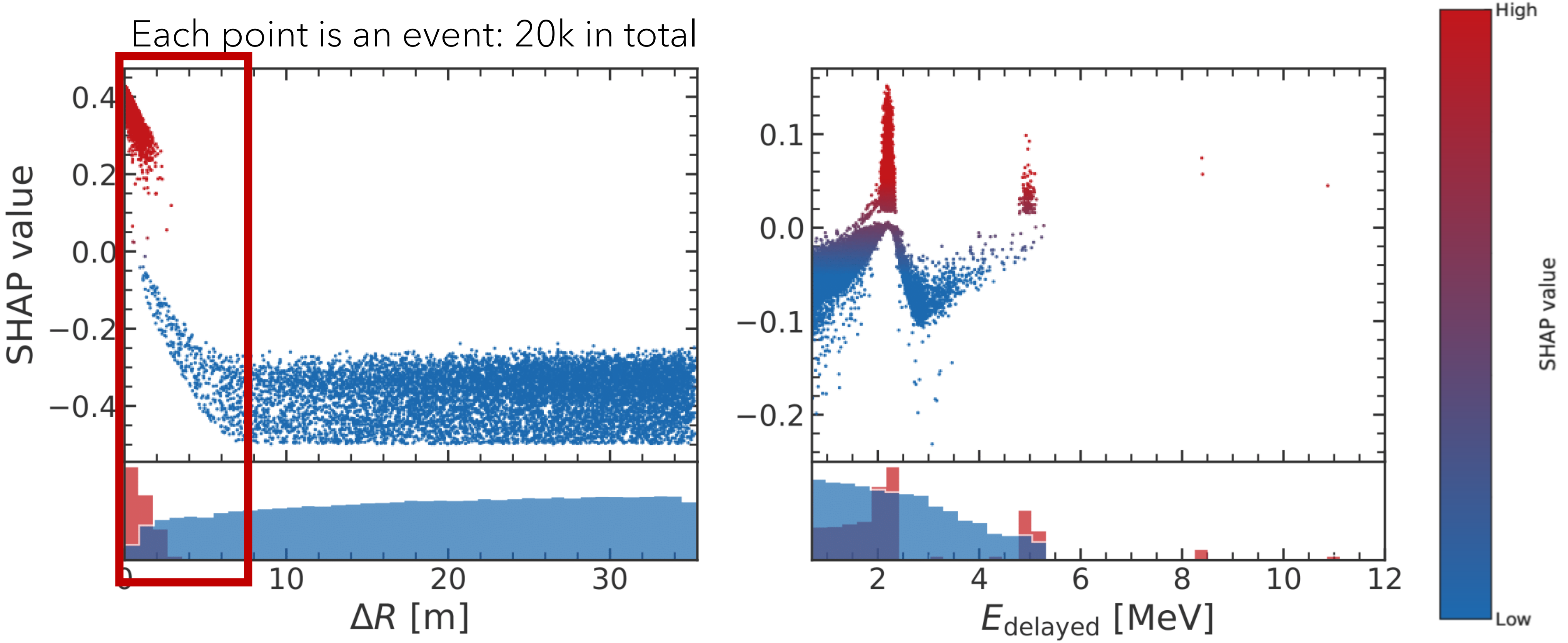
Table 1: Hyperparameter search space for FCNN. Selected hyperparameters are highlighted in bold.

NEURAL NETWORK: PERFORMANCE

Approach	*Volume*	*Efficiency*				
		0.2×Bkg	0.5×Bkg	1×Bkg	2×Bkg	5×Bkg
BDT	Full detector volume: $R < 17.7$ m	98.38%	98.81%	**99.02**%	99.19%	99.39%
	$R < 17.2$ m	91.58%	91.62%	**91.63**%	91.64%	91.64%
FCNN	Full detector volume: $R < 17.7$ m	96.94%	97.79%	**98.40**%	98.82%	99.21%
	$R < 17.2$ m	91.53%	91.60%	**91.63**%	91.64%	91.64%
Cuts	$R < 17.2$ m	—	—	89.90%	—	—

GLOBAL EXPLANATIONS FOR TOP FEATURES

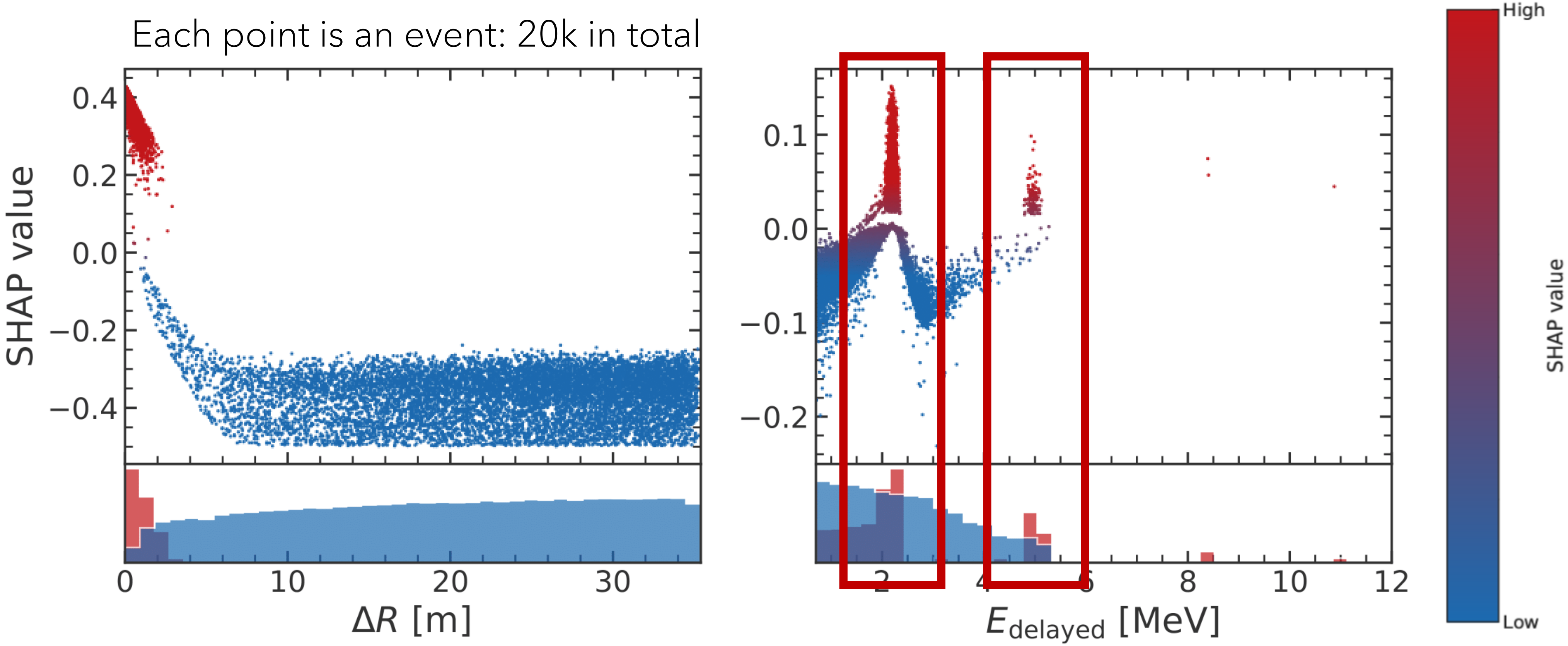
- Top-2 features distributions and SHAP values
- Small values of ΔR are very likely related to an IBD event



GLOBAL EXPLANATIONS FOR TOP FEATURES

- Top-2 features distributions and SHAP values
- Small values of ΔR are very likely related to an IBD event
- E_{delayed} has a clustered structure related to specific isotope on which **neutron** was captured

Each point is an event: 20k in total



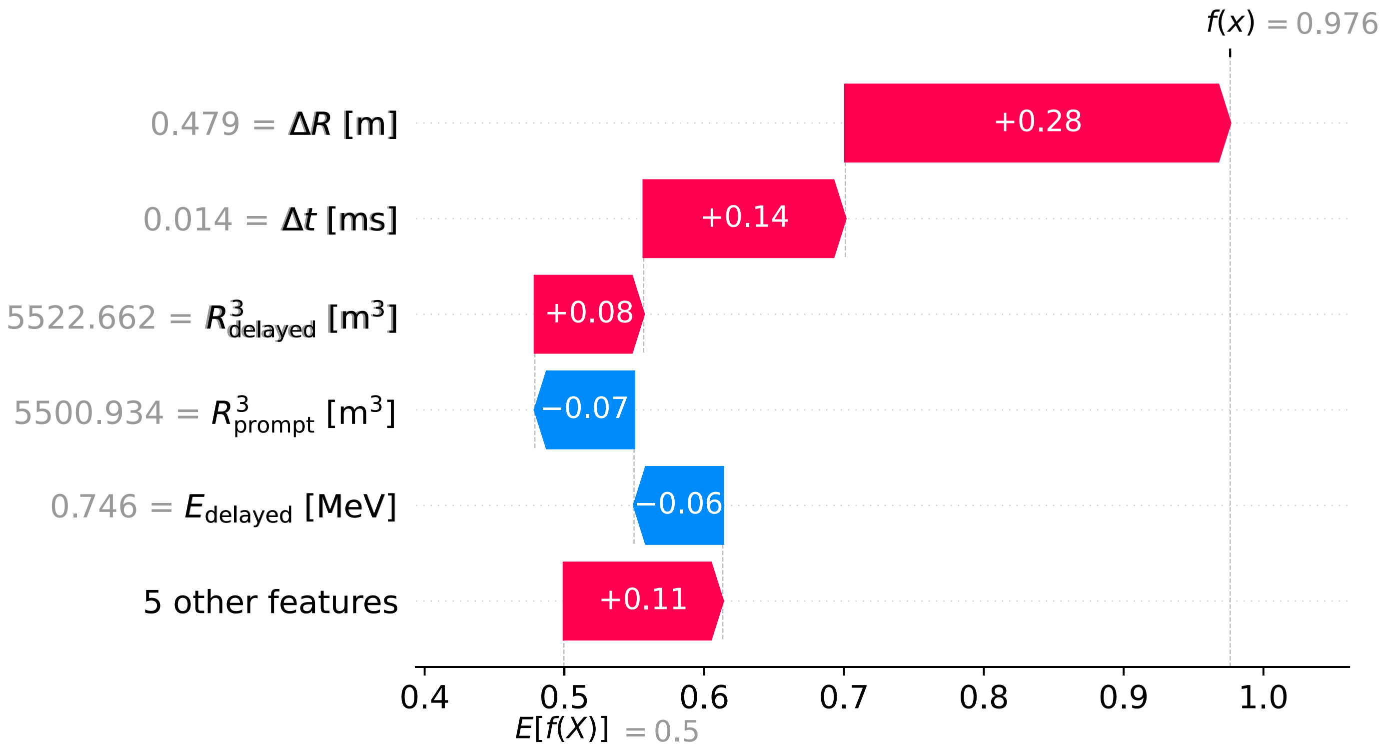
LOCAL EXPLANATIONS – GAMMA LEAKAGE

- Gamma produced by neutron capture escaping the target before depositing whole energy → gamma leakage

- Both prompt and delayed events occurred outside the FV

- This candidate pair would be **discarded** by cut-based selection → low E_{delayed} and FV cut

- NN classifies this as an IBD event based on the combination of other features → increase in efficiency ☺



INTERPRETABILITY AND SHAP VALUES

- SHAP (SHapley Additive exPlanations)
  - Calculate the contribution of each feature
  - Can be positive or negative
  - Show the **impact on predictions with respect to the average**

- Positive SHAP values → IBD class

- Negative SHAP values → accidentals class