Pile-up events Discrimination Based on Machine Learning in JUNO

Zhaoxiang Wu, Wenxing Fang, Guanda Gong, Miao He, Wuming Luo
wuzx@ihep.ac.cn

Aug . 23 , 2025

Institute of High Energy Physics , Beijing, China
University of Chinese Academy of Sciences, Beijing, China

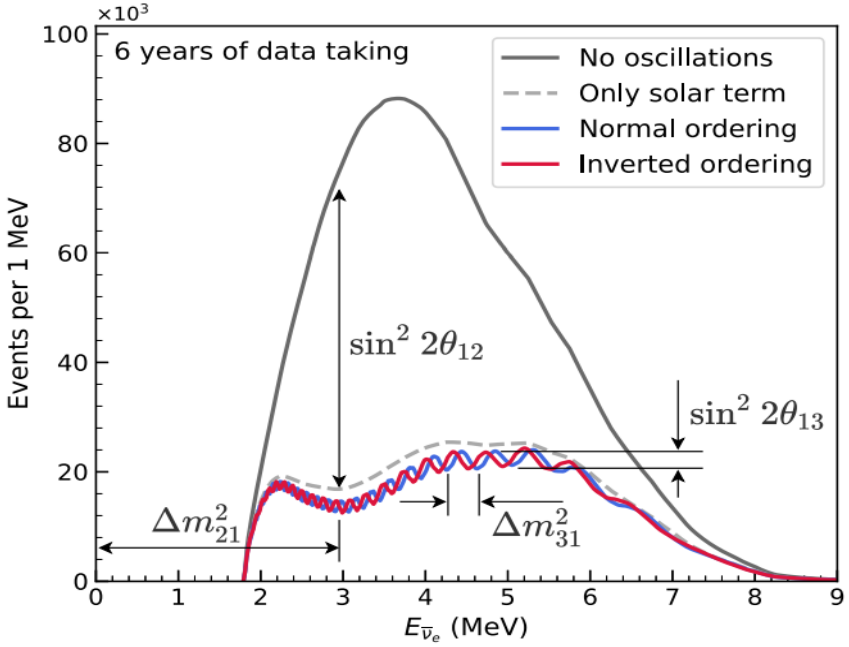
Outline

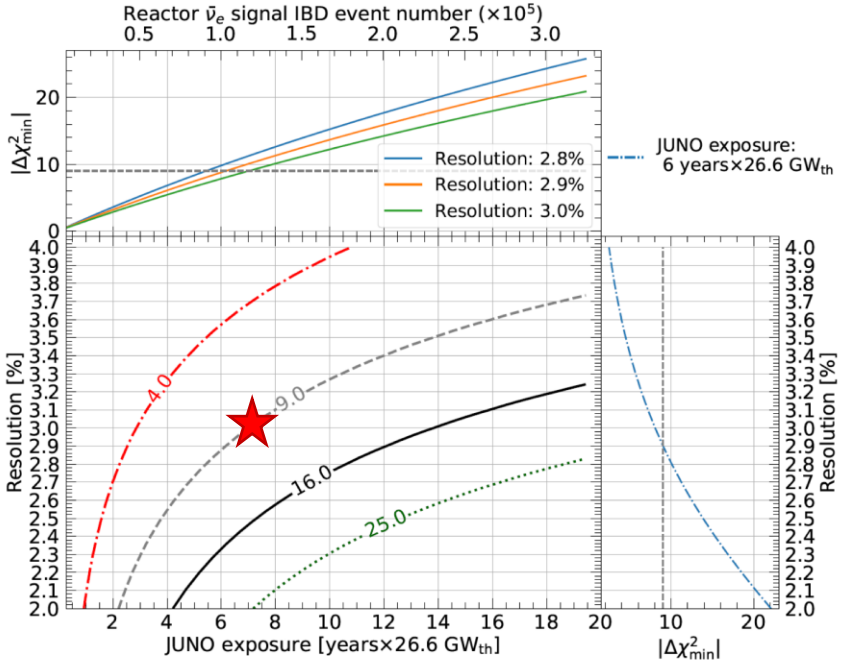
- **Introduction**
  - JUNO
  - Pile-up Events Characteristics
  - Impact of ^{14}C pile-up on Energy Resolution
- **Dataset**
- **Discrimination Based on Traditional Method**
- **Discrimination Based on Machine Learning**
  - 1D CNN
  - 1D Transformer
  - KamNet
- **Summary**

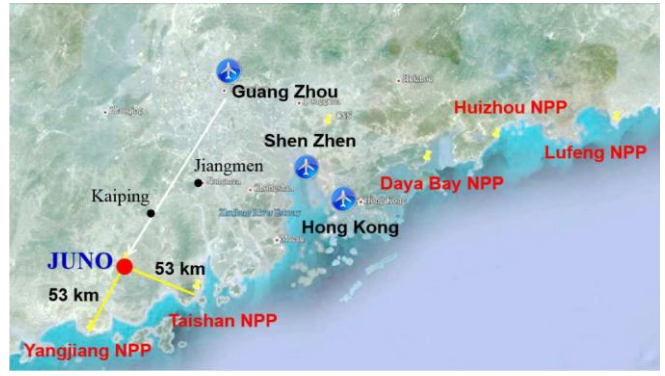
Introduction : JUNO

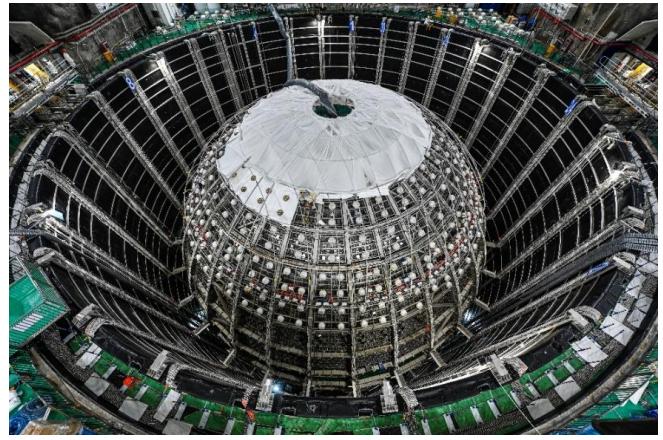
- JUNO starts data collection in 2025, its primary physics goal being to **determine the neutrino mass ordering** by measuring the reactor neutrino energy spectrum (IBD : $\bar{\nu}_e + p \rightarrow e^+ + n$)
- JUNO is 53 km away from both the Taishan and Yangjiang NPP
- Compared with other liquid scintillator (LS) detectors, JUNO has the **best energy resolution 3%@1MeV**, the **highest photocathode coverage**, and the **largest target volume**

实验	KamLAND	Borexino	Daya Bay	JUNO
光阴极覆盖率	34%	34%	12%	78%
靶质量	1000 ton	300 ton	20 ton	20000 ton
能量分辨率	6.5%	5%	8%	3%



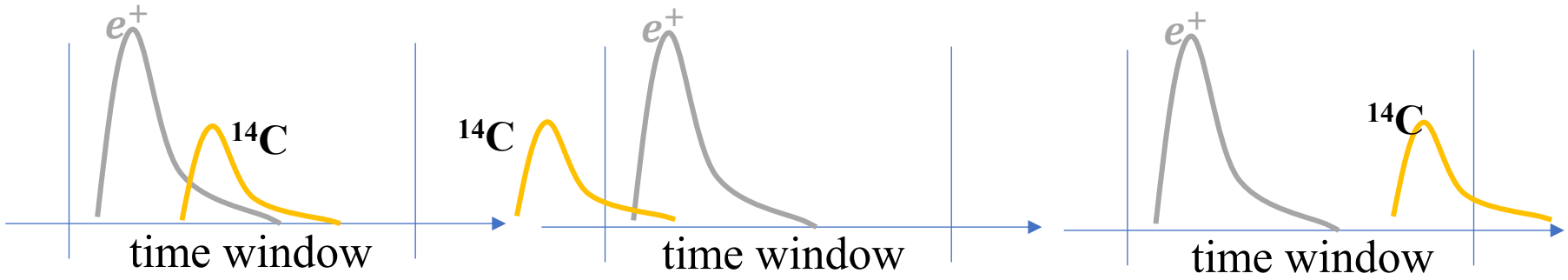


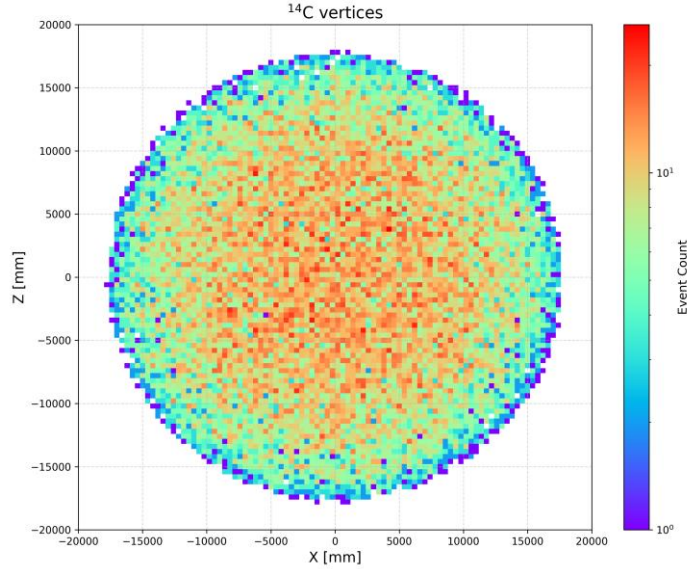




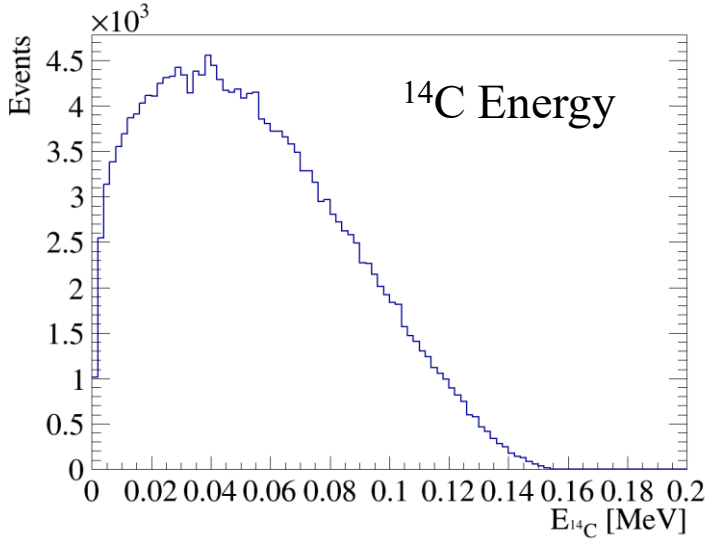
Introduction : Pile-up Events Characteristics

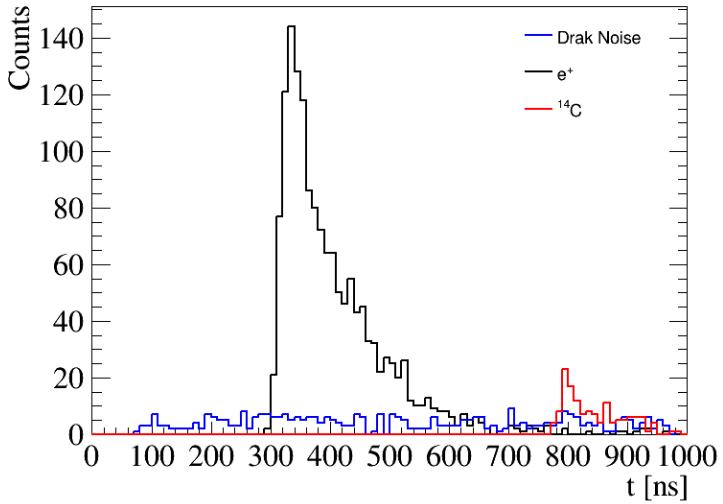
- ^{14}C β-decay **pile up** with e^+ signal in LS, forming pile-up events
  - ^{14}C and e^+ are **entirely uncorrelated**
  - **Random temporal** occurrence and **Spatially stochastic** distribution





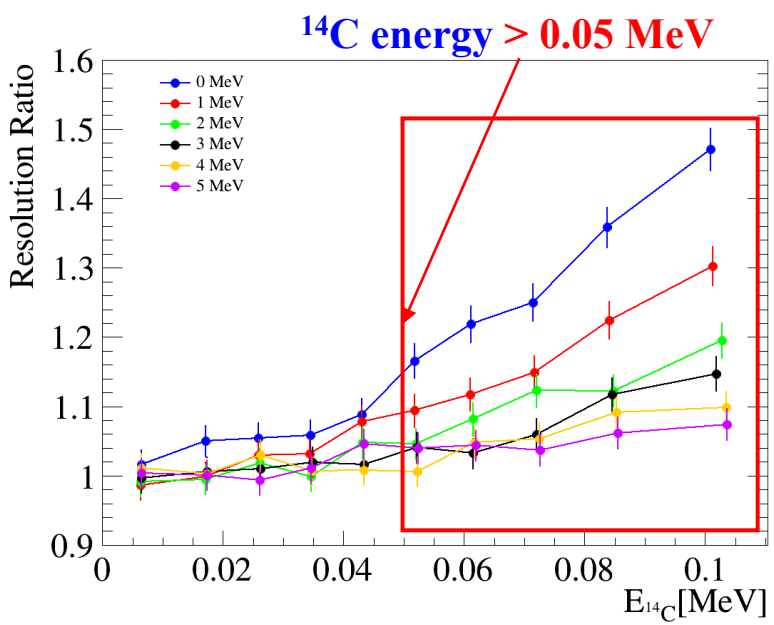
- **Distinct energy separation : e^+ deposit energy ≥ 1.022 MeV, ^{14}C β-decay yields energy < 0.16 MeV**

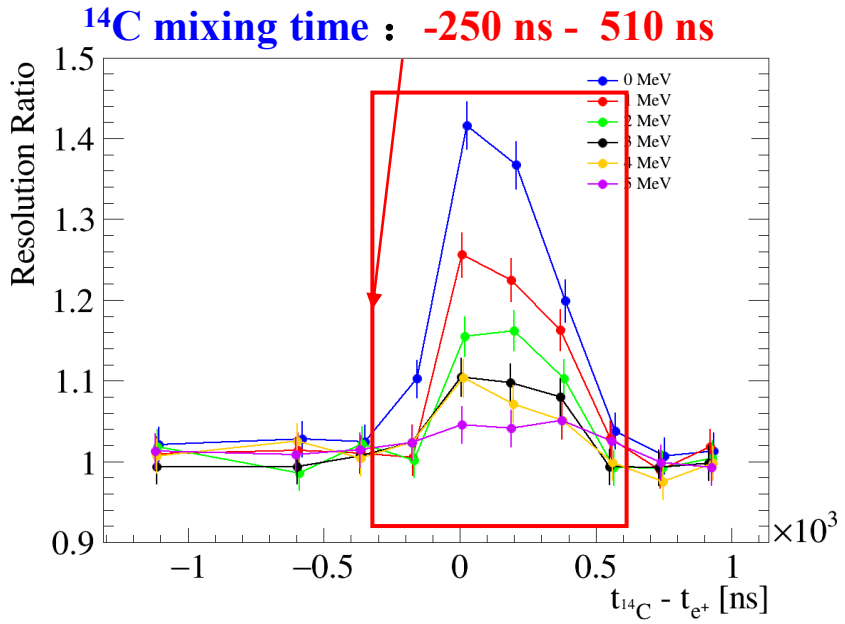


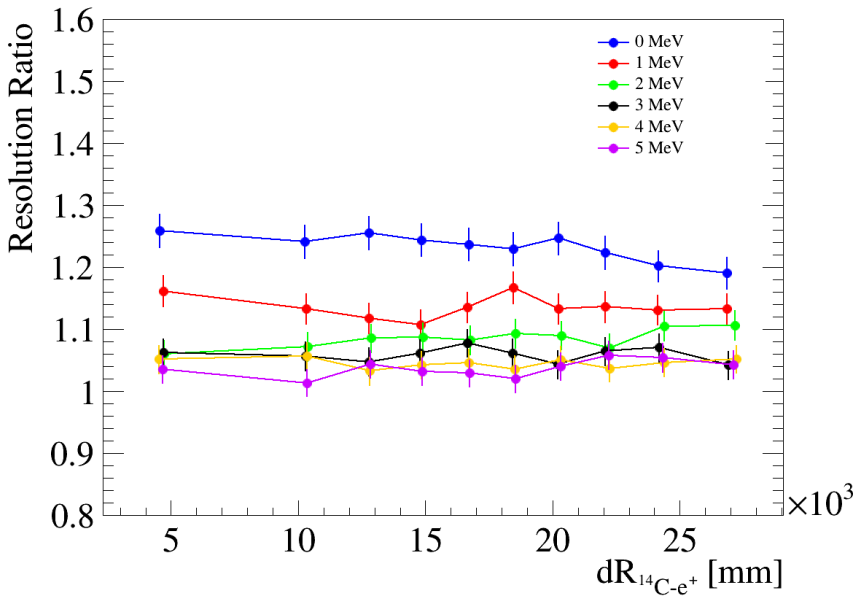


Introduction : Impact of ^{14}C pile-up on Energy Resolution

- Considering **^{14}C Energy**, **Mixing time** and **Distance with e^+**







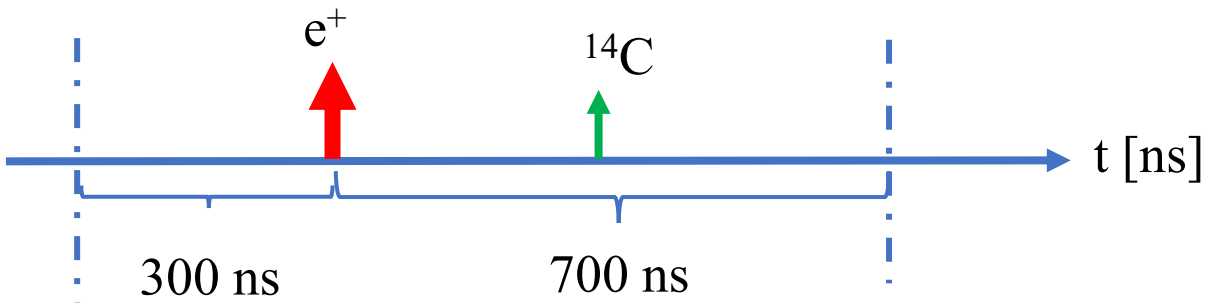
- About **1% ~ 2% ^{14}C** significantly affecting energy resolution
- Make discriminate e^+ and pile-up events be a significant challenges

Dataset

- **Training Sample :**

Training Sample	$t_{^{14}C} - t_{e^+}$ [ns]	Position	Energy [MeV]	Number [k]
e^+		uniform	0 ~ 5	400
Pile-up in window	[0, 200)	uniform	0 ~ 5	400
Pile-up out window	[-300, 0)∪[200, 700)	uniform	0 ~ 5	400

- **Test Sample :**
  - (0, 1, 2, 3, 4, 5) MeV e^+ or pile-up events : 500 k events in total / MeV
  - ^{14}C decay rate is set to 40 kHz

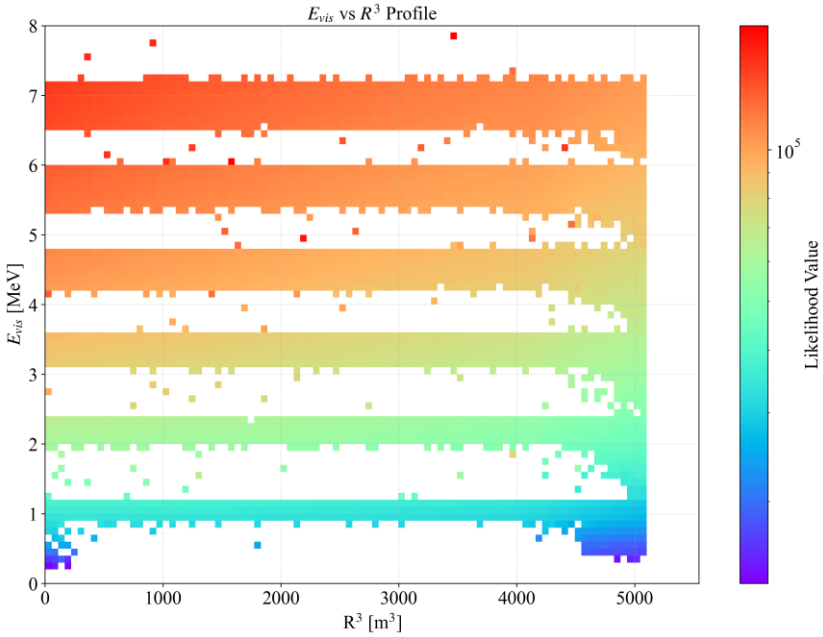


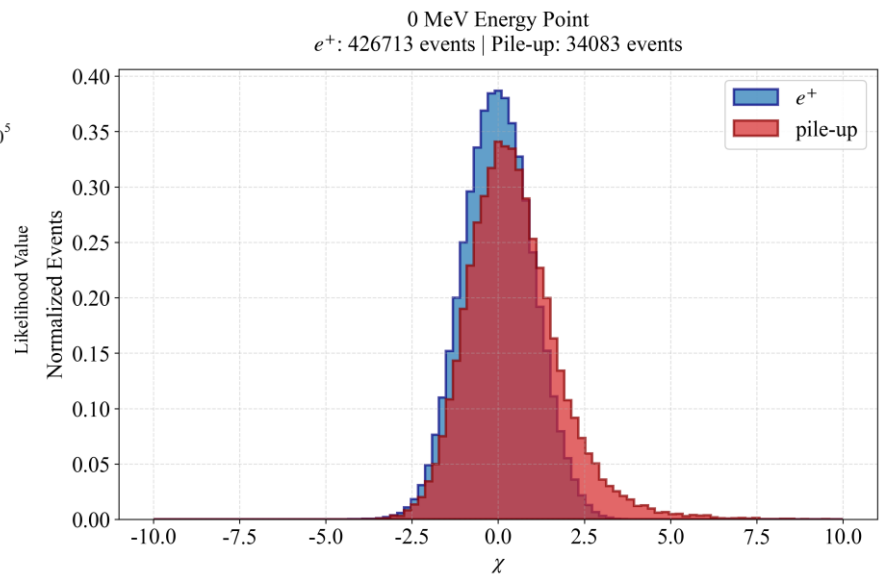
Discrimination Based on Traditional Method

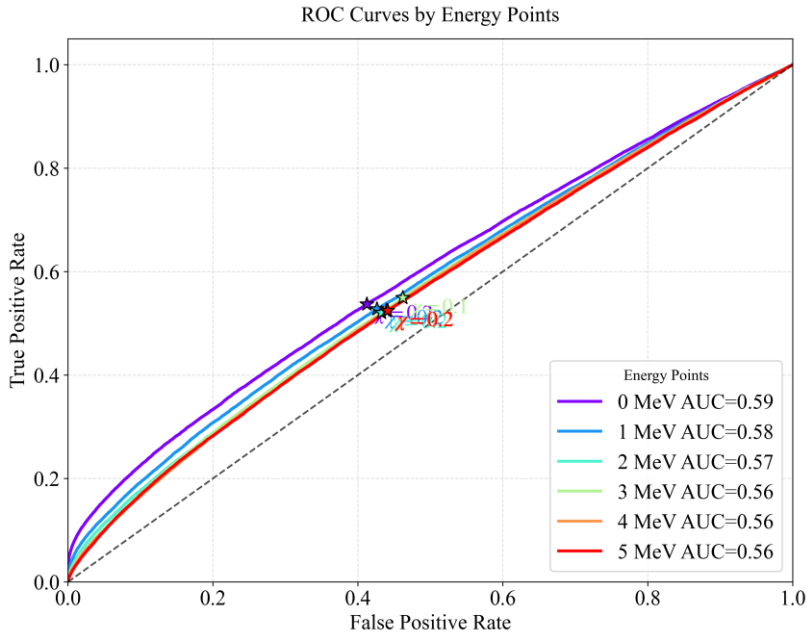
- JUNO reconstruction algorithm based on one event hypothesis, likelihood will be calculated during reconstruction
- We define a

$$\chi = \frac{L_{event} - L_{mean}(E_{vis}, R^3)}{L_{\mathrm{std}}(E_{vis}, R^3)}$$

- Pile-up events χ is larger than pure e^+



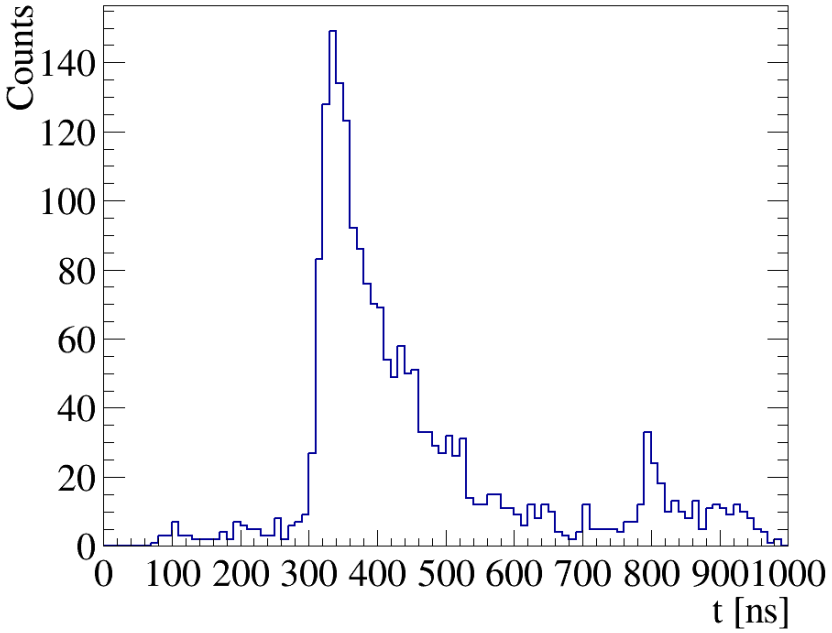


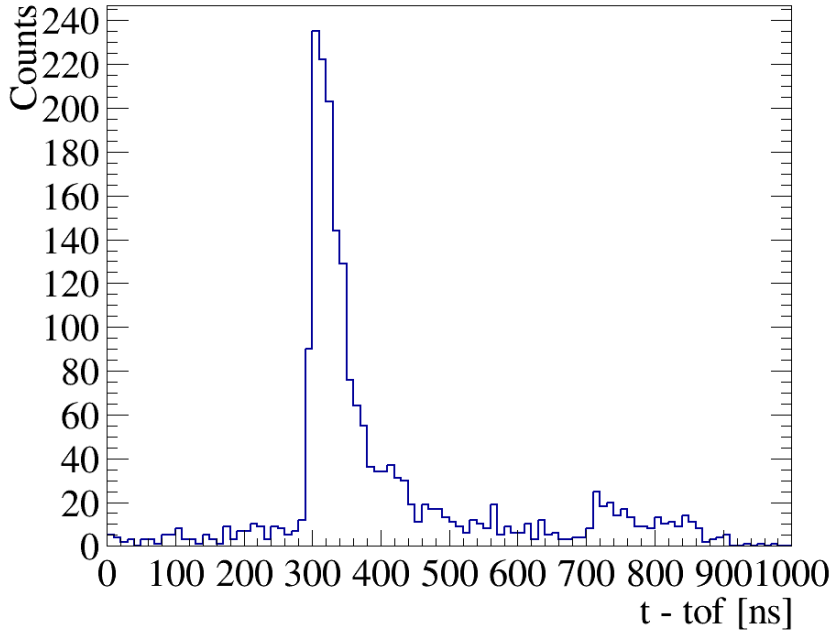


- Traditional methods are not effective.

Discrimination Based on Machine Learning : 1D CNN

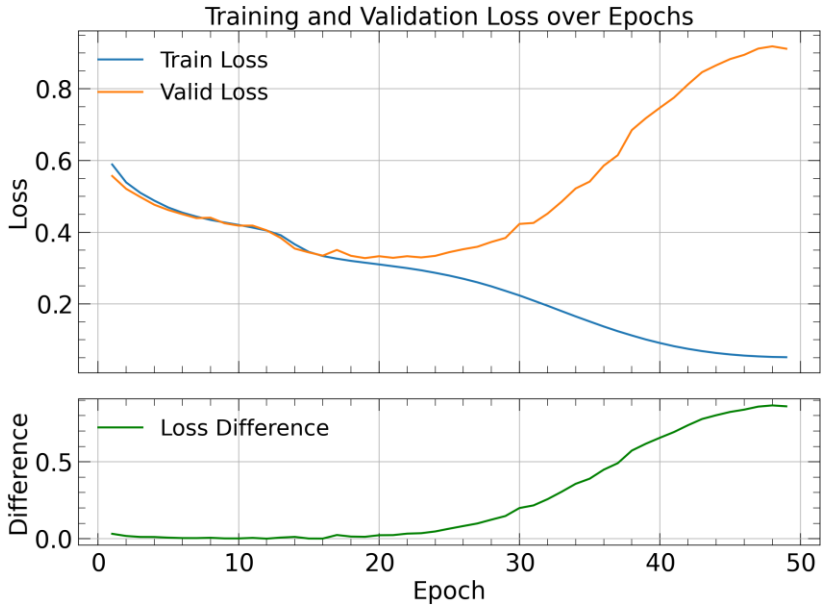
- Convolutional neural network
- Input
  - **hit time**
  - **hit time – tof (Time of Flight)**
- Output : probability of 2 categories

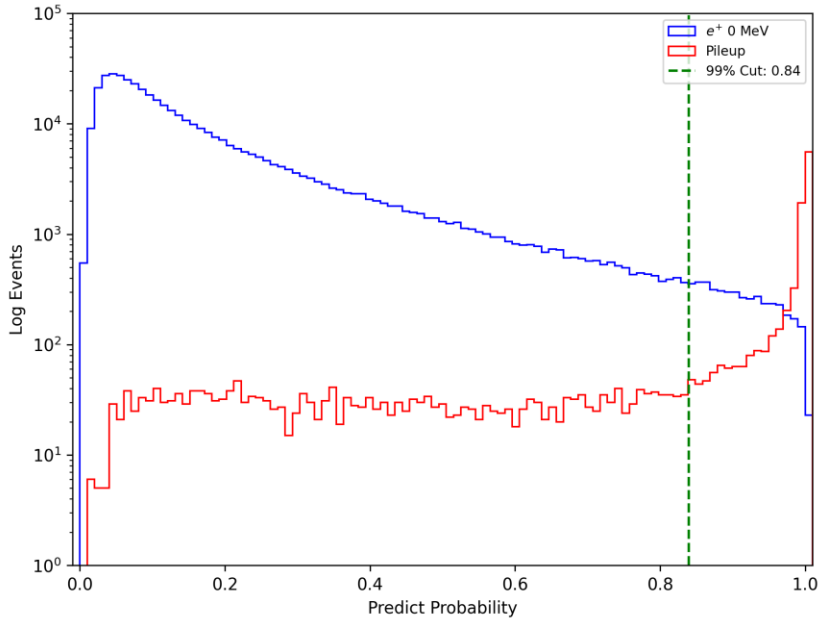


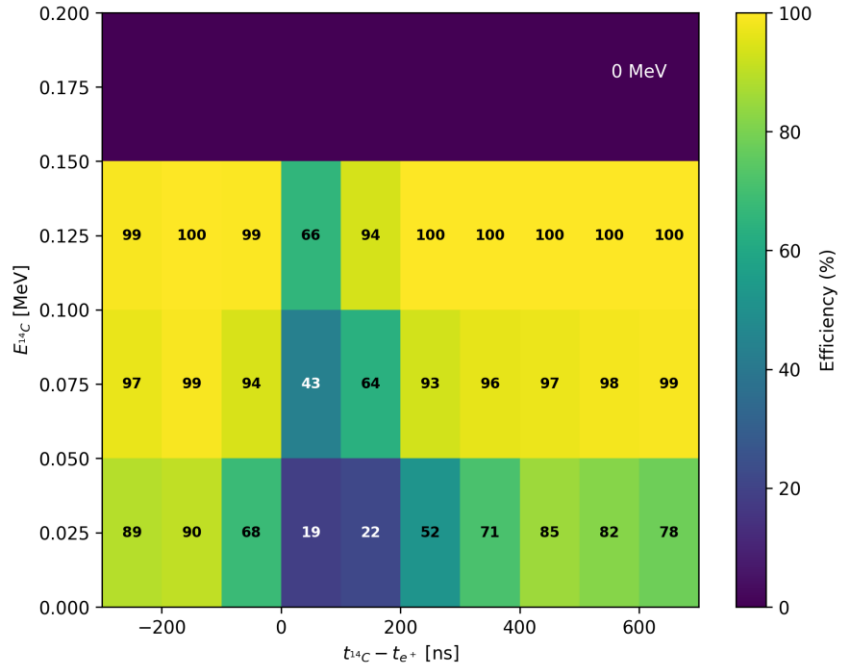


Input (2×1500)
Conv1d(2, 4, 3)
BatchNorm1d(4)
ReLU()+MaxPool1d(2,2)
Conv1d(4, 16, 3)
BatchNorm1d(16)
ReLU()+MaxPool1d(2,2)
Conv1d(16, 32, 3)
BatchNorm1d(32)
ReLU()+MaxPool1d(2,2)
Conv1d(32, 64, 3)
BatchNorm1d(32)+ReLU()
Conv1d(64, 64, 3)
BatchNorm1d(32)+ReLU()
MaxPool1d(2,2)
FC-1024
FC-128
FC-2

Discrimination Based on Machine Learning : 1D CNN



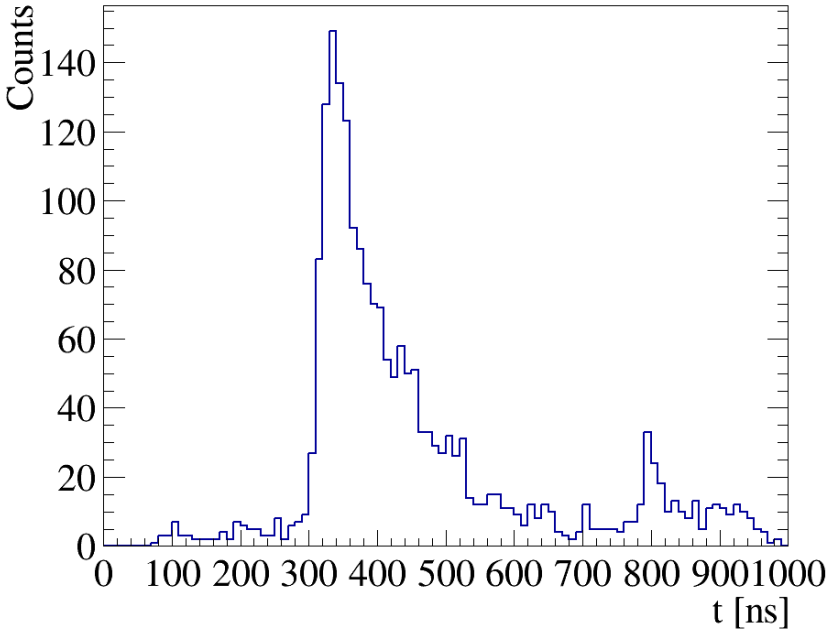


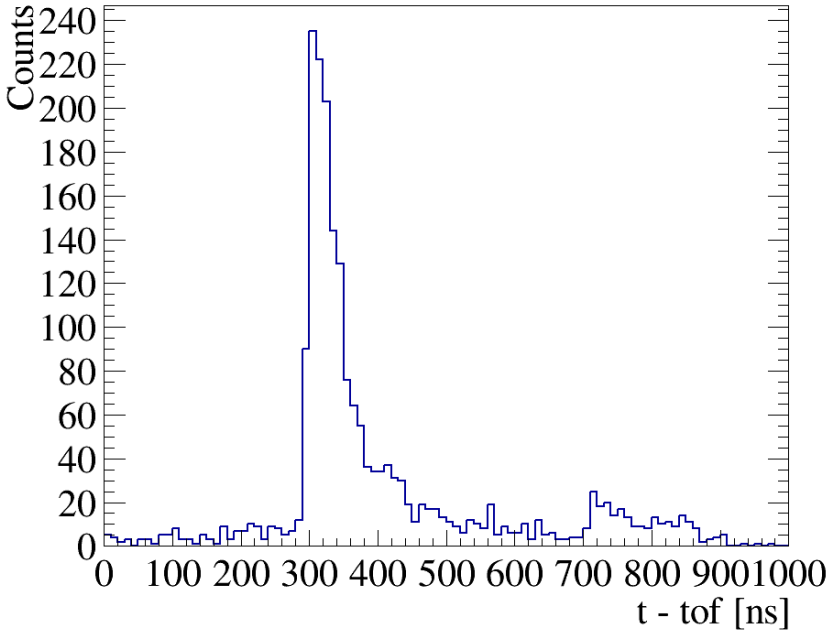


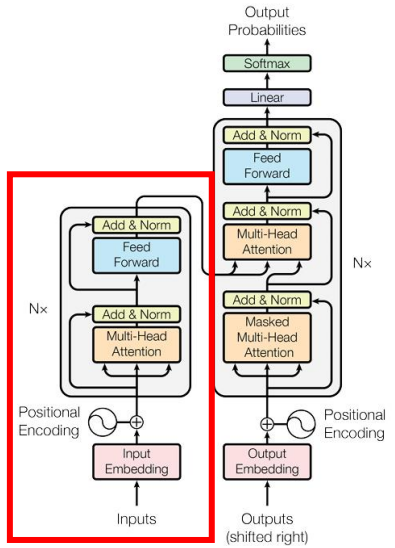
- **The most difficult discrimination region is when the mixing time is between 0 – 200 ns.**
- **^{14}C deposit energy > 0.05 MeV** and **mixing time between 0 – 200 ns**, the accuracy of CNN can reach **43% - 94%.**

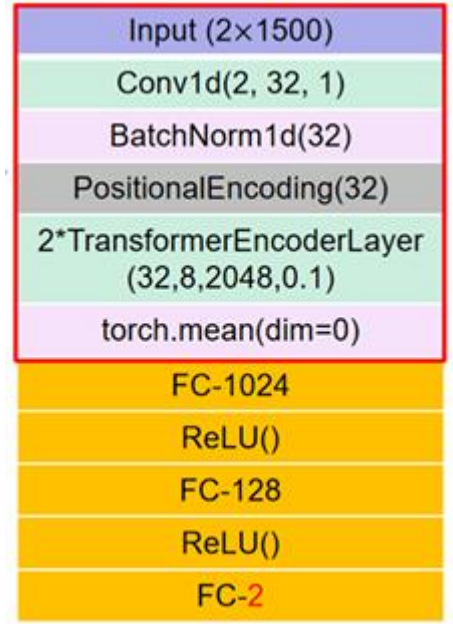
Discrimination Based on Machine Learning : 1D Transformer

- Transformer
- Input
  - **hit time**
  - **hit time – tof (Time of Flight)**
- Output : probability of 2 categories

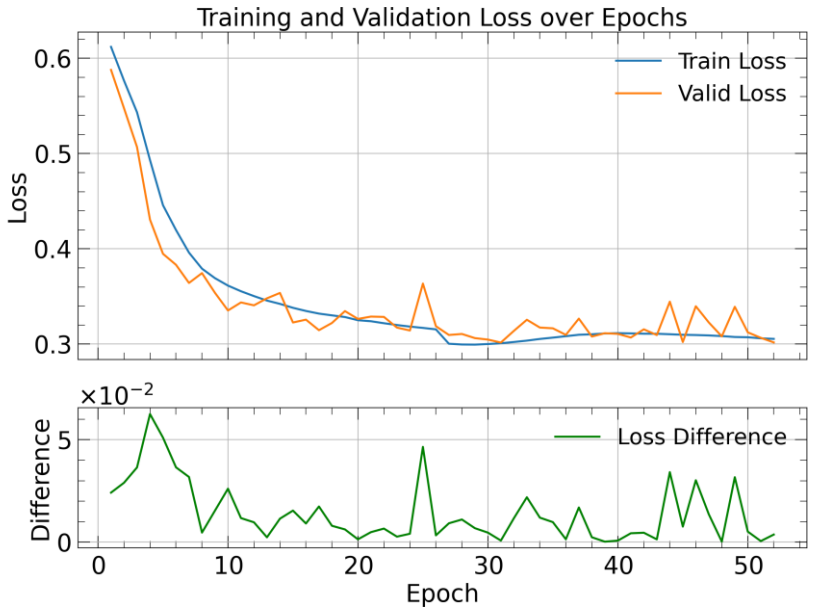


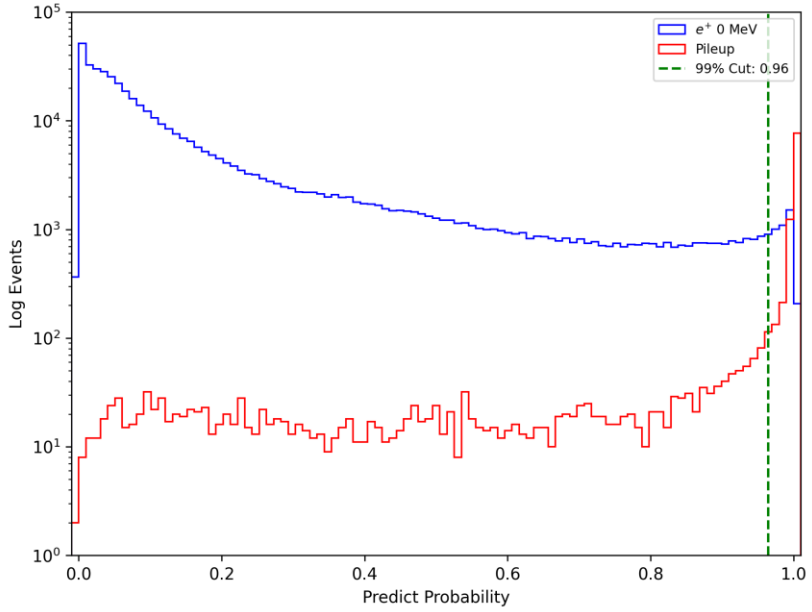


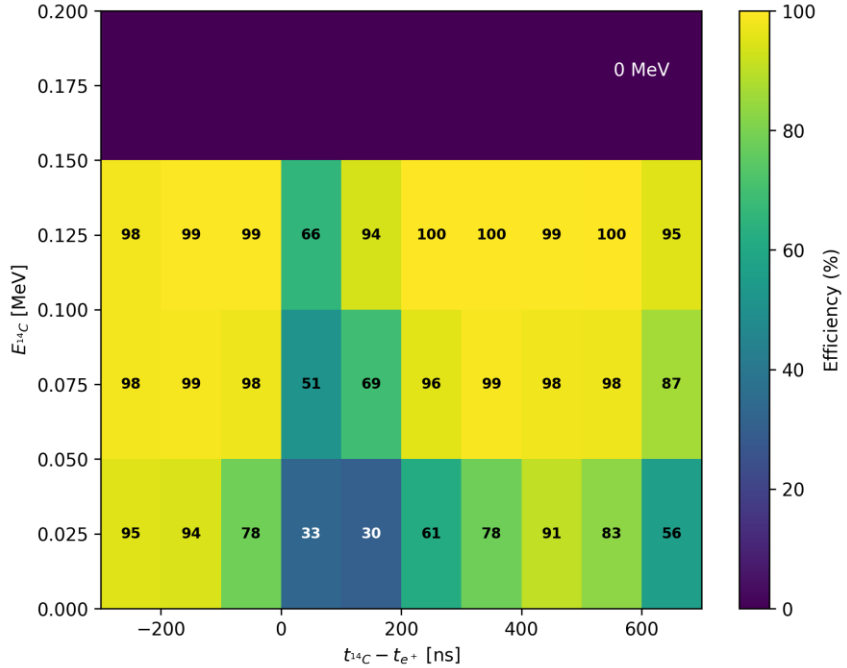




Discrimination Based on Machine Learning : 1D Transformer





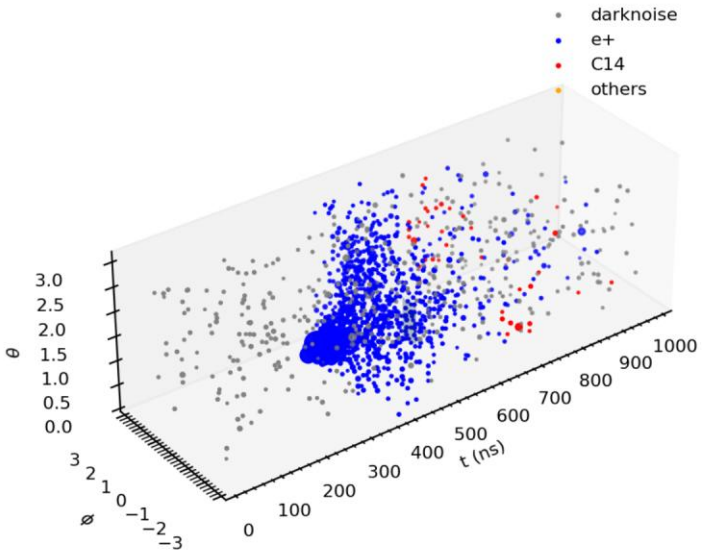


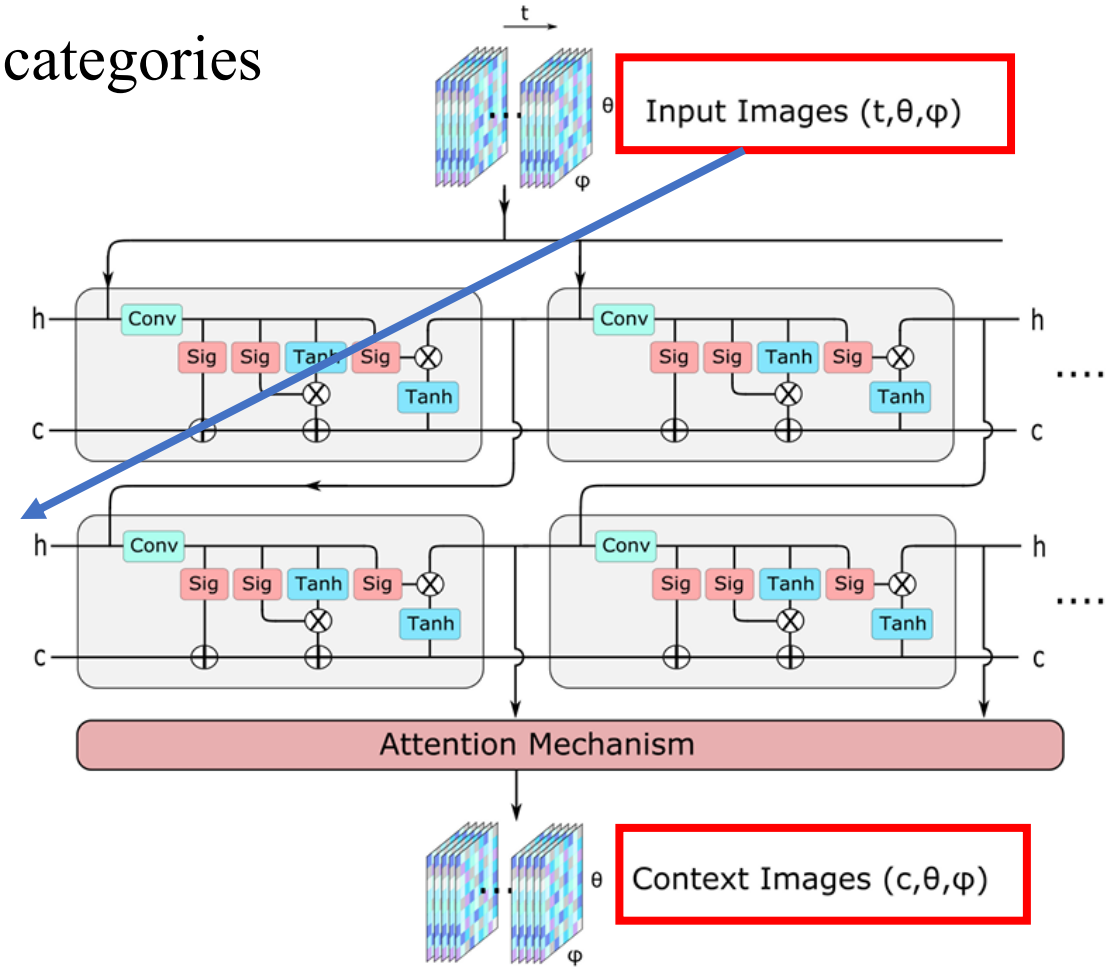
- **^{14}C deposit energy > 0.05 MeV** and **mixing time between 0 – 200 ns**, the accuracy of Transformer can reach **51% - 94%.**
- Transformer's performance is **better than CNN**

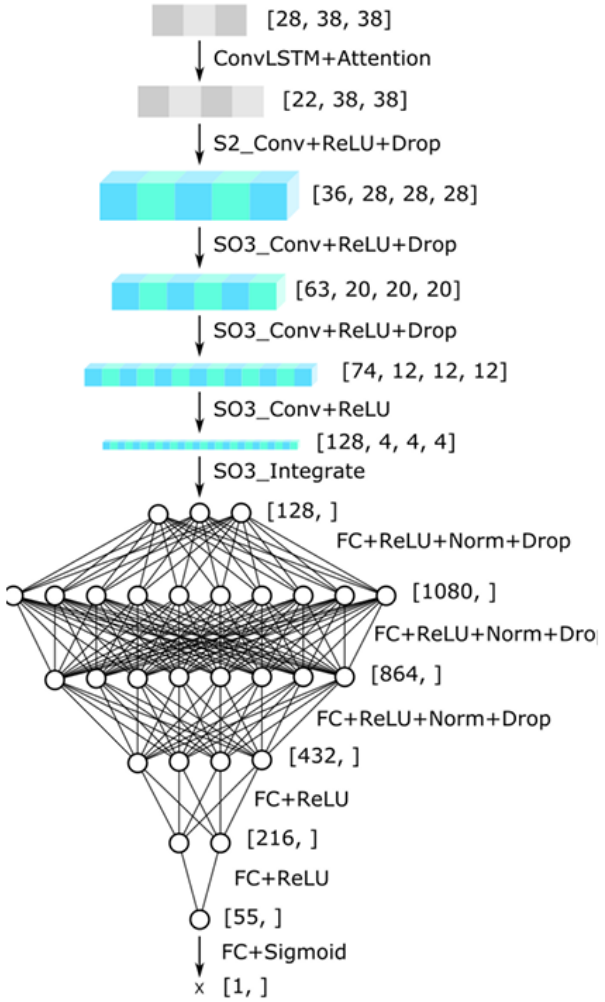
Discrimination Based on Machine Learning : KamNet

- KamNet : Sphere CNN + ConvLSTM + Attention
- Input : 3 dimension picture of (θ, φ, t)
- Output : probability of 2 categories

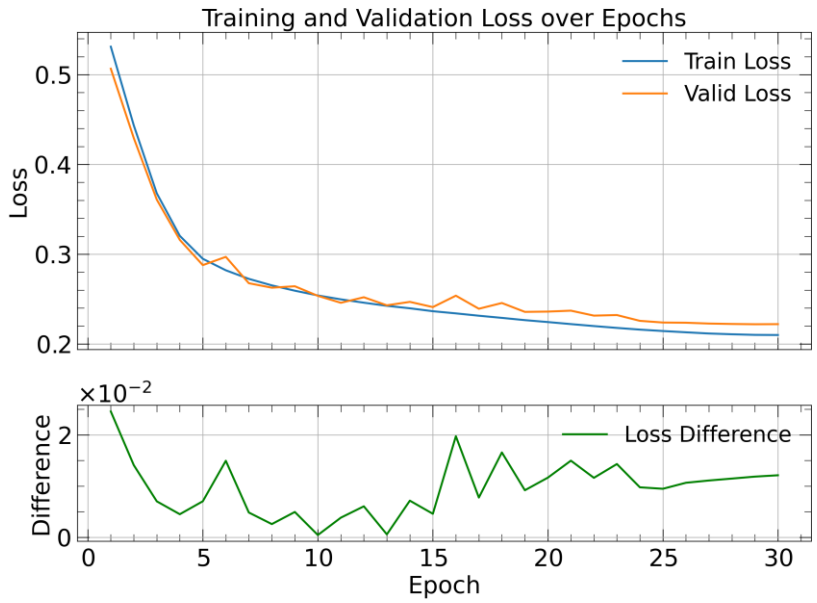
A.Li et al, Phys.Rev.C, 2023

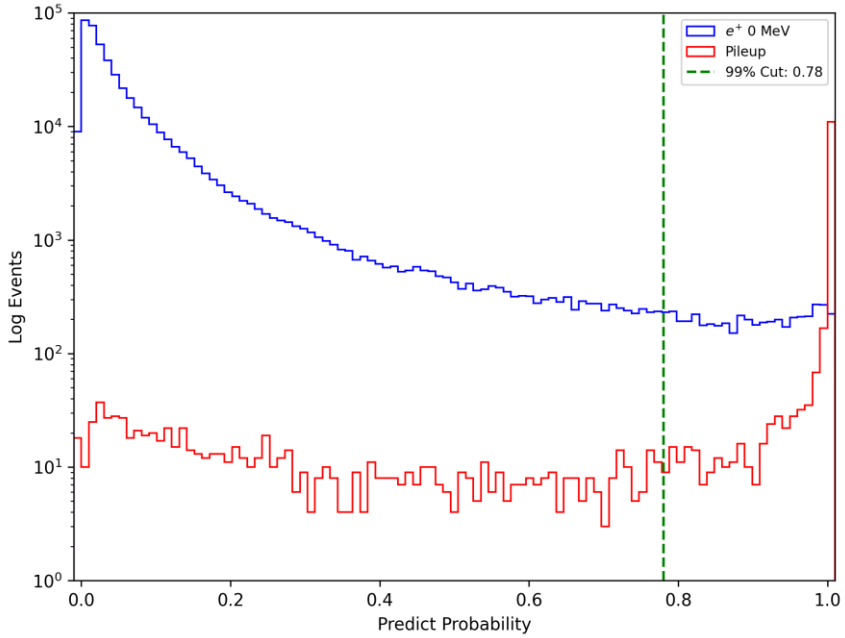


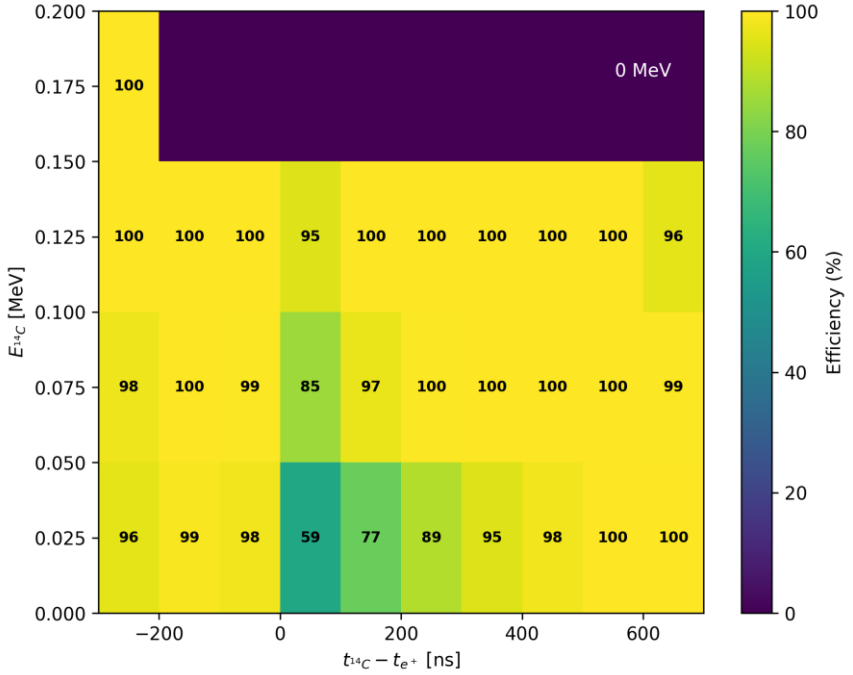




Discrimination Based on Machine Learning : KamNet







- **^{14}C deposit energy > 0.05 MeV** and **mixing time between 0 – 200 ns**, the accuracy of KamNet can reach **85% - 100%.**
- **KamNet's performance is best in these three models**

Summary

- **^{14}C deposit energy > 0.05 MeV** and **mixing time between 0 ~ 200 ns** is critical for events reconstruction and discrimination of pile-up events
- **CNN reach 43% - 94% accuracy** in the critical region
- **Transformer reach 51% - 94% accuracy** in the critical region
- **KamNet reach 85% - 100% accuracy** in the critical region, **with the best performance**
- Draft article has been completed.