IL NUOVO CIMENTO **47 C** (2024) 358
DOI 10.1393/ncc/i2024-24358-6

Colloquia: MAYORANA 2023

Energy reconstruction with machine learning techniques in JUNO

A. Gavrikov([1])([2]) on behalf of the JUNO Collaboration

([1]) *Dipartimento di Fisica e Astronomia dell'Università di Padova - Padova, Italia*
([2]) *INFN Sezione di Padova - Padova, Italia*

received 6 August 2024

Summary. — The Jiangmen Underground Neutrino Observatory (JUNO) is a multipurpose liquid scintillator neutrino experiment under construction located in China. Although the main source of neutrinos in JUNO is two nuclear power plants located about 52.5 km away from the experiment, it will also be able to study solar and atmospheric neutrinos, geoneutrinos and neutrinos coming from supernovae. The determination of the neutrino mass ordering (NMO) with 3-4σ in 6 years is the main goal of the experiment. Moreover, another JUNO's important aim is to measure neutrino oscillation parameters $\sin^2\theta_{12}$, Δm^2_{21}, Δm^2_{31} with sub-percent precision. The central detector of JUNO is an acrylic sphere filled with 20 kt of liquid-scintillator (LS) surrounded by 17612 20-inch photomultiplier tubes (PMTs) and 25600 3-inch PMTs, providing $\sim$78% coverage of the detector sphere. Thanks to the almost complete coverage of the sphere by the PMTs array, as well as high light yield leads to an unprecedented, for LS-based experiments, energy resolution of 3% at 1 MeV. Due to the need to take into account various effects, including the non-linearity of the energy response and the detector's spatial non-uniformity, event energy reconstruction is not a straightforward task. In this study, energy reconstruction for reactor neutrino events with machine learning (ML) techniques is presented. The following two models are used: Boosted Decision Trees and Fully Connected Deep Neural Network. The models are trained on aggregated features extracted from charge and time information on PMTs.

1. – Introduction

Liquid-scintillator (LS) detectors surrounded by photomultiplier tubes (PMTs) are widely used for studying neutrino nature in many modern neutrino experiments [1]. The goal of the new generation of LS-based experiments is the construction of even larger and more precise apparatuses. One example of such experiments is the Jiangmen Underground Neutrino Observatory (JUNO). JUNO is a neutrino observatory with a broad physics program located in China about 52.5 km away from two power plants: Taishan and Yangjiang [2], in a laboratory submerged 650 meters deep underground. The JUNO's central detector (CD) is an acrylic sphere, 35.4 meters in diameter, filled with 20 kt of LS. The CD is held by a stainless steel construction immersed in an ultra-pure water pool. To collect photons emitted in LS, the CD of JUNO is equipped with two types of PMTs: 17612 large 20-inch PMTs and 25600 small 3-inch PMTs. This PMT array covers almost 78% of the detector sphere, leading to a high photoelectron statistic of $\sim$1600 photoelectron per 1 MeV. Figure 1 shows a schematic view of JUNO.

JUNO has a broad physics program and the main goal is to determine the neutrino mass ordering (NMO) within 3-4σ in 6 years of data-taking. Furthermore, JUNO will measure the following oscillation parameters $\sin^2\theta_{12}$, Δm^2_{21}, Δm^2_{31} with sub-percent precision.

The "golden reaction" for detecting neutrinos – Inverse Beta Decay (IBD) – is used in JUNO to detect reactor antineutrinos: $\overline{\nu}_e + p \rightarrow e^+ + n$. The positron rapidly deposits its kinetic energy and annihilates into two 0.511 MeV gammas forming a prompt signal. Thereafter, the neutron, after approximately $\sim$200 μs, is captured on hydrogen (99%) or carbon (1%) emitting a 2.22 MeV or 4.95 MeV gamma, respectively. Such a clear signature of two signals allows an effective selection of IBD events and reject backgrounds. Most of the neutrino energy converts into positron energy. Therefore, the energy of the positron can be reconstructed and the initial neutrino energy recovered using the following expression: $E_{\tilde{\nu}_e} \approx E_{e^+} + 0.8$ MeV. In this study, we apply energy reconstruction for separated positron events.

The IBD's positron (and its secondary electrons) produces scintillation photons and Cherenkov light in the target. Hereafter, the part of the produced light is absorbed by the PMT array and brings information about the positron's energy and as a consequence, the initial energy of the neutrino. JUNO's large PMTs have a quantum efficiency of approximately 30%, therefore a similar fraction of the optical photons that reach a PMT will eject photoelectrons. After a waveform reconstruction procedure, for each fired PMT, two variables are stored: charge and first hit time (FHT). The amount of charge collected, temporal distribution and spatial distribution of the signal are used to reconstruct the deposited energy of an event.

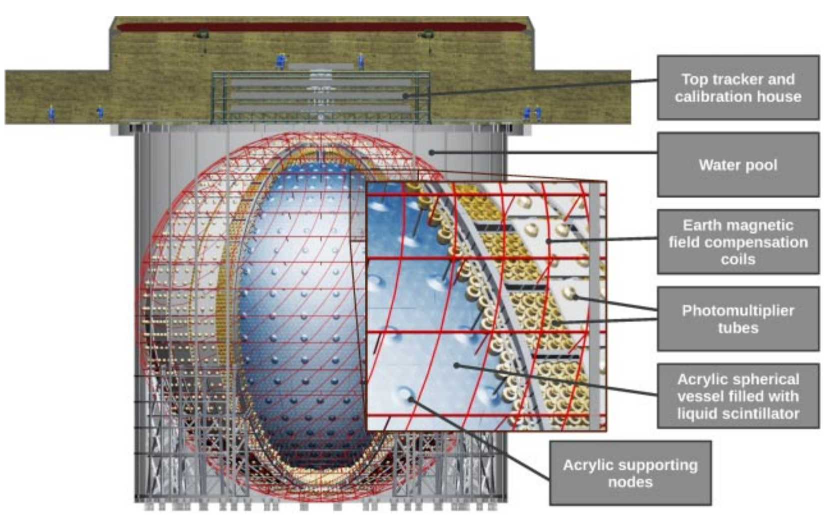

Fig. 1. – Schematic view of the JUNO central detector and other main components

2. – Machine learning approach

In this study, only large (20”) PMTs are used, since they provide 96.5% of the photo-covered area of the detector. Thus, each event is described by 35224 variables: by charge and FHT values at each PMT (the values might be zeros if a PMT is unfired). To reduce the dimensionality of the input, we perform an additional data processing procedure. We use aggregated information extracted from charge and FHT distributions as input features to ML models. The aggregation is driven by the physics of the event and exploratory data analysis. Boosted Decision Trees (BDT) and Fully Connected Deep Neural Network (FCDNN) approaches were chosen as powerful tools to process tabular data. The study is a continuation of the one proposed in ref. [4] and uses an updated JUNO software [5]. ML approaches that handle PMT-wise information, instead of aggregated features, can be found in ref. [6]. In addition to ML studies, the classical method based on PMT-wise inputs can be found in ref. [7].

2·1. *Feature engineering*. – The constructed features are divided into two categories: charge-related features and time-related ones. Both categories (charge and time) provide independent information about the signal. Table I collects and briefly describes the full set of engineered features and the detailed description can be found in ref. [4]. The most important feature is `AccumCharge`: the total charge collected on all fired PMTs. Other features, provide additional necessary information about the signal to account for different effects. The most essential ones are energy non-linearity and detector non-uniformity. For example, the center of charge and center of FHT provides a rough approximation of the vertex and help to correct the detector response non-uniformity. To provide more precise information about the signal, moments and percentiles of these distributions are used. The total feature set contains 91 features.

To train and test models, we prepared the corresponding datasets: the training one and the testing one. Both of them are generated by the full detector Monte Carlo method using the official JUNO software [5]. The training dataset consists of ∼2.25 million positron events uniformly spread in the CD with a flat kinetic energy spectrum $E_{\mathrm{kin}} \in [0, 10]$ MeV. The testing dataset consists of 14 subsets (50k events each) with discrete kinetic energies uniformly distributed in the detector.

Many of the 91 features, described above, bring similar information about the signal and we want to optimise it to keep a subset which provides the same performance of the

Table I. – *List of all feature notations with brief descriptions. Here,* $n \in \{2, 5, 10, 15, \ldots, 90, 95\}$.

Notation	*Description*
`AccumCharge`	Total accumulated charge
`nPMTs`	Number of fired PMTs
x_{cc}, y_{cc}, z_{cc}, R_{cc}, θ_{cc}, ϕ_{cc}, J_{cc}, ρ_{cc}, γ_z^{cc}, γ_y^{cc}, γ_x^{cc}	Center of charge
$\mathrm{pe}_{\mathrm{n}\%}$, $\mathrm{pe}_{\mathrm{mean}}$, $\mathrm{pe}_{\mathrm{std}}$, $\mathrm{pe}_{\mathrm{skew}}$, $\mathrm{pe}_{\mathrm{kurtosis}}$	Charge distribuion
x_{cht}, y_{cht}, z_{cht}, R_{cht}, θ_{cht}, ϕ_{cht}, J_{cht}, ρ_{cht}, γ_z^{cht}, γ_y^{cht}, γ_x^{cht}	Center of FHT
$\mathrm{ht}_{\mathrm{n}\%}$, $\mathrm{ht}_{\mathrm{n_{i+1}\%-n_i\%}}$, $\mathrm{ht}_{\mathrm{mean}}$, $\mathrm{ht}_{\mathrm{std}}$, $\mathrm{ht}_{\mathrm{skew}}$, $\mathrm{ht}_{\mathrm{kurtosis}}$	FHT distribution

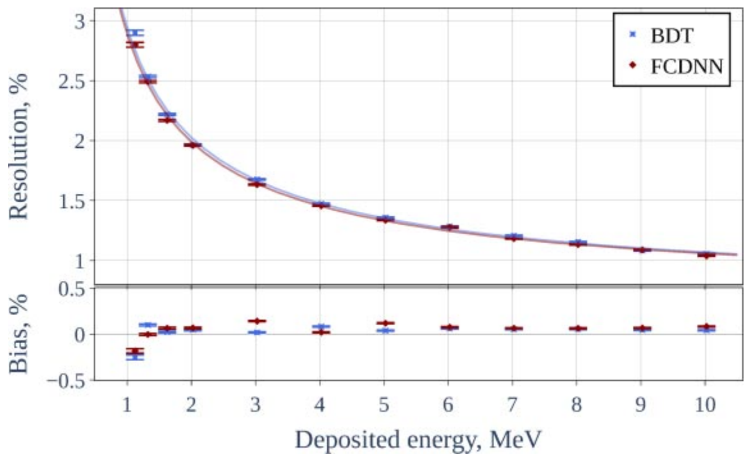

Fig. 2. – Energy reconstruction performance for BDT and FCDNN. The upper panel shows resolution and the lower panel represents bias.

model as the total set. By using a greedy algorithm for feature selection described in ref. [4], we kept the following 15 features (sorted by importance):

$$\texttt{AccumCharge}, R_{\rm cht}, {\rm J_{cc}}, {\rm ht}_{20\%-15\%}, {\rm pe}_{\rm std}, {\rm nPMTs}, {\rm z_{cc}}, {\rm ht}_{\rm std}, {\rm ht}_{30\%-25\%},$$
$$R_{\rm cc}, {\rm ht}_{5\%-2\%}, {\rm pe}_{\rm mean}, {\rm ht}_{25\%-20\%}, {\rm ht}_{10\%-5\%}, {\rm ht}_{35\%-30\%}.$$

3. – Results and conclusions

Figure 2 shows the energy reconstruction performance (resolution and bias) of the BDT and FCDNN models. The hyperparameter optimization of the models is performed and follows the same procedure as in ref. [4].

This study presents machine learning approaches for energy reconstruction of positron events in the JUNO detector. We process signals from each large PMT: the charge collected by it and the first hit time, computing aggregated features. Thus, the input dimensionality reduces from 35224 to 15, which makes it possible to train efficient and fast models to deal with tabular data: BDT and FDNNN.

* * *

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.

REFERENCES

[1] Sajjad M. *et al.*, *Prog. Part. Nucl. Phys.*, **124** (2022) 103947.
[2] JUNO Collaboration (Abusleme A. *et al.*), *Prog. Part. Nucl. Phys.*, **123** (2022) 103927.
[3] JUNO Collaboration (Abusleme A. *et al.*), *Chin. Phys. C*, **46** (2022) 123001.
[4] Gavrikov A. *et al.*,, *Eur. Phys. J. C*, **82** (2022) 1021.
[5] Lin T. *et al.*,, *Eur. Phys. J. C*, **83** (2023) 382.
[6] Qian Z. *et al.*,, *Nucl. Instrum. Methods Phys. A*, **1010** (2021) 165527.
[7] Huang G. *et al.*, *Nucl. Sci. Tech.*, **34** (2023) 83.