

Performance of machine learning algorithm for predicting muon multiplicity in the INO-ICAL prototype

Lakshmi Murgod,^{a,*} Deepak Samuel^a and Sourav Dutta^a

^aCentral University of Karnataka, Department of Physics,
Kalaburagi-585367, India
E-mail: lakshnimurgod@cuk.ac.in , deepaksamuel@cuk.ac.in,
duttasourav65925@gmail.com

In most of the analyses using the cosmic muon data from the INO-ICAL prototype stack, only single muon events are considered. Multi-muon events appear to be noisy events to the algorithm and thus get rejected, reducing the physics potential of the detector. To address this issue, we have developed an ML algorithm to predict muon multiplicity in cosmic muon event registered in the INO-ICAL detector at TIFR. In this work, we present the performance of this algorithm in terms of its efficiency to correctly tag multi-muon events and also to predict the multiplicity.

37th International Cosmic Ray Conference (ICRC 2021)
July 12th – 23rd, 2021
Online – Berlin, Germany

*Presenter

1. Introduction

The India-based Neutrino Observatory (INO) collaboration has proposed to build a 50 kton magnetized Iron CALorimeter (ICAL) detector with a rock overburden of about 1 km. About 30,000 Resistive Plate Chambers (RPCs) will be used as position-sensitive detectors each of about 200 cm x 200 cm in area [1]. The ICAL detector will be sensitive to muons and will look for precision measurements of neutrino oscillation parameters and determine the neutrino mass ordering. As part of detector design, development, and characterization studies for INO, many prototype RPC stacks were built. These stacks have been used for many detector studies and, more recently, for cosmic muon studies. Here we present our first simulation results of predicting cosmic muon multiplicity using machine learning for the prototype stack at TIFR.

2. Prototype stack at TIFR

The prototype stack consists of 12 layers of RPCs separated by a distance of 16.8 cm. The RPCs used in this setup consist of two parallel electrodes of a glass of dimensions 1 m x 1 m x 3 mm, separated by a gap of about 2 mm and across which a voltage of 10 kV is applied. The gap is filled with Freon, Isobutane, and SF₆ in the ratio 95.5:4.2:0.3. The RPCs in the prototype stacks are all operated in the "avalanche" mode. On top and bottom of the electrodes, pick-up panels containing 32 strips, each of 28 mm width, made of copper are placed. The strips in the top pick-up panel run orthogonal to the strips in the bottom pick-up panel. These strips are connected to front-end electronics for signal processing and readout. This arrangement enables the readout of the x-y interaction point of a particle passing through the detector along with the time of interaction. The spatial resolution of the RPCs in the prototype is about 3 cm and the timing resolution is in the order of 1 ns. The block diagram of the TIFR prototype stack is shown in figure 1

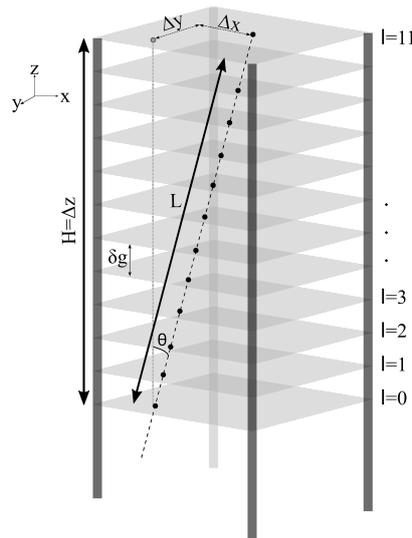


Figure 1: Schematic of the prototype detector stack at TIFR [2].

3. Datasets

A python-based algorithm has been developed to generate the dataset for training the machine learning models. To quantify and understand the effect of detector parameters like detector efficiency (η), strip hit multiplicity ($S_m = 1.5$) and noise multiplicity ($N_m = 3$), four different datasets were generated and analyzed which are described in the next sections.

3.1 Dataset I: Clean tracks

Cosmic muon tracks have been generated using straight-line equation 1, for randomly generated slope m and intercept c .

$$s_x = m_x l_z + c_x \quad (1)$$

Where s_x or s_y is the strip hit from the x or y side respectively for z^{th} layer of the stack 1. From the above equation, if we know the slope and intercept, the strip number corresponding to a given layer number can be estimated. Tracks were generated for both xz and yz planes and events to be used in the analysis were required to have the same number of hits in both x and y projections. Only events in which at least one track has crossed certain minimum layers (i.e, trigger condition is added) in both the projection are considered. Every event in the dataset is represented by a sparse matrix of dimension 12×64 corresponding to 12 layers and 32 strips in the two projections with each strip in a layer represented by an element in the matrix. The matrix is initialized to zero and then elements corresponding to strips fired in an event are set to 1.

For multi-muon events, a random number generator is used to generate multiple random slopes and intercepts for a given event. The dataset is generated such that there is a uniform distribution of events amongst different multiplicities.

3.2 Dataset II: Tracks including detector efficiency

To simulate a realistic detector, the detector efficiency η should also be considered while generating the datasets. Dataset II represents such events in which this factor is considered. We have assumed a detector efficiency of 80% for events in this class. To generate these events, a uniform random number $D1$ is generated between 0 and 100 for each hit in a track. If the generated random number $D1$ is less than η then the strip hit is removed from the track [3].

3.3 Dataset III: Tracks including η and strip hit multiplicity (S_m)

Another important parameter to be considered to simulate a realistic detector is the average number of strips fired for a minimum ionizing particle hit, called cluster size or strip multiplicity. High cluster size worsens the accuracy of the particle track localization. From previous studies, the average strip hit multiplicity of the RPCs for cosmic ray muon data is found to be ~ 1.5 [4]. Events in the class of Dataset III represent tracks with strip hit multiplicity and detector efficiency factored in.

To generate tracks including strip hit multiplicity effect, a uniform random number $S1$ is generated between 0 and 150 for every track. If generated number $S1$ is between 0 and 50, one extra hit is added on the left side of the main strip. If $S1$ is greater than 100 hits is added on the right side of the main strip and if the generated random number is greater than 50 and less than 100, the hits corresponding to these numbers are allowed to remain the same. [3].

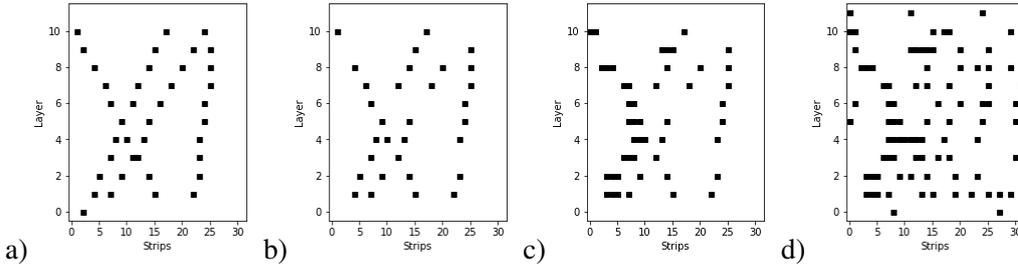


Figure 2: Cosmic muon event from each dataset. a) Dataset I, b) Dataset II, c) Dataset III, and d) Dataset IV

ML Model	Number of events for each multiplicity	Maximum multiplicity	Total number of events
I	1000	11	11000
II	1000	7	7000
III	1000	5	5000
IV	1000	3	3000

Table 1: The four ML models and the number of events used for training and testing the model.

3.4 Dataset IV: Tracks including η and S_m and noise multiplicity (N_m)

In addition to the detector efficiency and strip hit multiplicity, there is a probability that hits are also generated due to detector noise like improper gas distribution or electronic noise. This effect causes hits to be formed away from the muon track and is quantified by the noise multiplicity parameter. To account for this parameter, extra hits were added to each layer by generating a random integer number R1 for each layer between 0 and N_m . Then, R1 random integers between 0 and 32 are generated and the corresponding strip numbers are set to be hits [3]. In this study the maximum noise Considered is $N_m=3$ hits per layer.

Figure 2 describes the generation of a multiplicity 4 event with 9 minimum layers and for each of the datasets mentioned in subsection 3.

4. Training the Machine learning (ML) models

For this study, we used sci-kit-learn, an open-source python-based package [5]. Specifically, the XGBoost Classifier model was used to predict the cosmic muon multiplicity. The four ML models were developed using the XGBoost algorithm tabulated in the table 1, based on four different datasets described in section 3. 80% of data was used for training the model and the remaining 20%, used for testing the model.

5. Results

To study the performance of the models developed, the multiplicity prediction efficiency was analyzed, the results of which are shown in figure 3. For clean tracks, an efficiency of 80% is

ML Model	Time taken for training (min)	Time taken for testing (sec)
I	4.2	0.4
II	3	0.3
III	1.5	0.2
IV	0.3	0.03

Table 2: Time taken for data generation, training and testing each model.

observed even for a multiplicity 11 event. For dataset IV, the efficiency rapidly decreases to 80% even for multiplicity 3 events. Nevertheless, such events are expected to be very rare in real data.

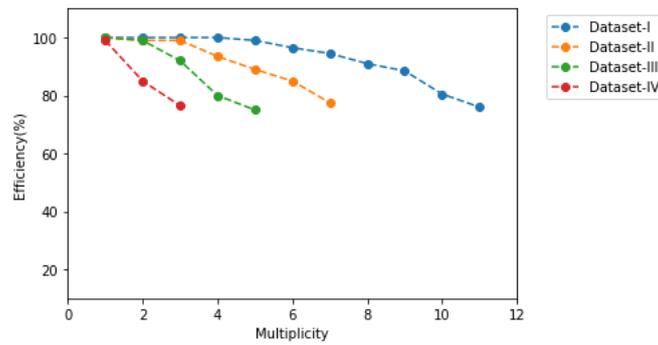


Figure 3: Efficiency plot as a function of multiplicity for all four datasets

All the datasets were trained and tested on a system with windows 10 Pro, a 64-bit operating system, and an Intel Core i7-7500U CPU with 16 GB physical memory. The time taken for training and testing the model is described in table 2.

6. Conclusions

We have presented our first results on the performance of a machine learning model to predict muon multiplicity using a machine learning model using a simulated dataset. We have used the INO-ICAL prototype detector setup as our geometry in the simulation. We report a high prediction efficiency of 80% for muon multiplicity of 11 without noise. In a real case, due to the quality of the detectors in terms of the noise and efficiency, we expect tracks very similar to the clean tracks. However, we also notice a degradation in the prediction efficiency when extreme detector conditions are assumed (i.e detector efficiency 80%). The prediction is fast and therefore the trained model can be implemented in the DAQ code to tag multi-muons directly while writing to disc. We intend to use this model to filter out multi-muon events from real data and carry out further analysis.

References

- [1] Santonico, R., and R. Cardarelli. *Development of resistive plate counters.*, *Nuclear Instruments and Methods in physics research* **187.2-3** (1981) 377-380.

- [2] Samuel, Deepak, Pratibha B. Onikeri, and Lakshmi P. Murgod., *Angular resolution of stacked resistive plate chambers.*, *Journal of Cosmology and Astroparticle Physics* **2017.01** (2017) 058.
- [3] Samuel, Deepak, and Karthik Suresh., *Artificial neural networks-based track fitting of cosmic muons through stacked resistive plate chambers.*, *Journal of Instrumentation* **13.10** (2018) P10035.
- [4] Bheesette, Satyanarayana., *Design and characterisation studies of resistive plate chambers.* (2009).
- [5] Raschka, Sebastian., *Python machine learning.*, Packt Publishing Ltd, (2015).