SUPERVISED EVENT CLASSIFICATION IN AN OPTICAL TIME PROJECTION CHAMBER*

V. Guadilla, N. Sokołowska, M. Pfützner

Faculty of Physics, University of Warsaw, 02-093 Warsaw, Poland

Received 30 November 2022, accepted 2 January 2023, published online 22 March 2023

Machine Learning algorithms trained on Monte Carlo simulated data are proposed for the classification of experimental data from an Optical Time Projection Chamber. In this contribution, we describe the simulation procedure to mimic experimental data, as well as the algorithms chosen for nuclear physics cases of β-delayed (multi)-particle emission and two-proton radioactivity. A proof of principle of the whole procedure is discussed for the decay of ^{11}Be.

DOI:10.5506/APhysPolBSupp.16.4-A37

1. Introduction

Time Projection Chambers (TPC) provide a wide range of applications in nuclear physics. One of them is the study of the radioactive decay of exotic nuclei emitting charged particles. Triggered by the discovery of two-proton radioactivity [1, 2], a powerful technique for such studies was developed at the University of Warsaw based on the use of a TPC with optical read-out (OTPC) [3, 4]. In gas mixtures based on Ar, He, and small amounts of CF_4, the visible light produced in the amplification stage of the detector, composed of several gas electron multipliers, is registered by a CCD camera and a photomultiplier tube (PMT). The combination of this information allows for the full 3D reconstruction of the tracks of the charged particles emitted by the nuclei [5].

The search for rare events in decay experiments with a TPC is many times hampered by the difficulties in classifying the registered signals. The identification of rare decay branches of the order of 10^{-3}–10^{-6} among the huge amount of data typically recorded may be like looking for a needle in a haystack. In addition, sometimes the fingerprint of such an exotic branch happens to be very similar to some background or to other decay branches, which causes difficulties in standard classification and analysis of data.

* Presented at the Zakopane Conference on Nuclear Physics, *Extremes of the Nuclear Landscape*, Zakopane, Poland, 28 August–4 September, 2022.

Machine Learning (ML) techniques are increasingly used in nuclear and particle physics, and they have already proved their applicability to other TPC set-ups, such as the Active-Target Time Projection Chamber (AT-TPC) [6, 7]. In the present work, we use supervised algorithms based on Monte Carlo (MC) simulations to classify OTPC events. The OTPC signals used for training the ML models are simulated by means of the Geant4 simulation package [8], as will be detailed in the next section. Geant4 simulations have already been employed to train supervised classification algorithms in other nuclear physics setups (see Refs. [9, 10] for recent examples). MC simulations are a convenient way of training ML models in situations where there is a lack of labeled experimental data or the available statistics for the different classes of events is very unbalanced, as in β-delayed multi-particle emission or two-proton radioactivity studies, due to the small branching ratios associated with these events.

2. Monte Carlo simulations

Geant4 is a simulation toolkit for particle transport and interaction between matter and radiation. The energy loss of charged particles is simulated based on condensed history models, which rely on the evaluation of the energy loss in steps, based on stopping power data. Computationally this is more efficient than simulating every single interaction, an approach that, in addition, suffers from the scarcity of cross-section data required for all the materials involved. In this work, we use the version 10.5 of Geant4 and the energy loss (for non-relativistic charged particles, in our case) is evaluated by the processes G4hIonisation (for protons) and G4ionIonisation (for generic ions and α particles) by means of the models G4BraggModel and G4BraggIonModel, respectively. For user-defined materials, as it is in our case, both models employ the ICRU49 parameterisation [11]. The final energy loss is computed after considering fluctuations, in order to take into account the straggling effect associated with the stochastic character of the process.

For the simulation of OTPC events, the proper gas composition is defined, with the associated temperature and pressure conditions. The electric field inside the chamber volume is considered as well. The simulated energy loss is stored together with spatial information about the point where it was deposited. The spatial coordinates are defined as a random position between the start and the end of the computation step (of the order of 1 μm). Experimentally calculated spatial resolutions are applied, with different values for x–y (corresponding to the CCD camera) and z (corresponding to the PMT). The experimental pixel size of the CCD and the binning of the PMT signal are taken into account to project the energy loss information and spatial coordinates into 16-bit unsigned PNG images (x–y) and one-dimensional

arrays (z). Experimental noise distributions are sampled to add this effect to both output files, thus mimicking the experimental data features. An illustration of the energy loss simulation and projection outputs is shown in Fig. 1.

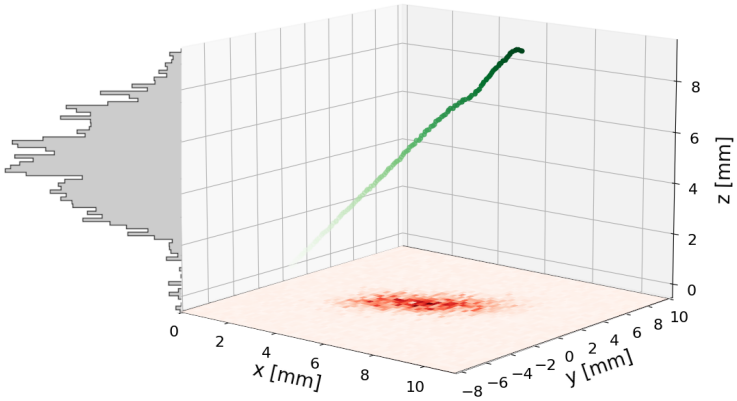

Fig. 1. Example of energy loss simulation in 3D with the corresponding projections in x–y and z, mimicking the CCD image and the PMT signal, respectively. Experimental spatial resolutions and noise distributions are taken into consideration.

3. Machine Learning approach

The most natural way of applying ML methods with images relies on the use of Convolutional Neural Networks (CNN), as in previous approaches with other TPCs [6]. However, such algorithms are computationally very demanding and typically complex predefined models for pattern recognition are used. In our approach instead of using images, the three projections (in x, y, and z) from CCD and PMT are used. In fact, the full projections are not needed, and we just consider some parameters that help to describe and characterize such projections, as the length, the mean, the kurtosis or the skewness. This allows us to adopt computationally simpler models. In particular we use the very efficient extreme gradient boosting algorithm (XGboost) as well as Neural Networks (NN) algorithms, based on the implementations in the Keras deep learning library written in Python [12]. In both cases, standard ML techniques like k-fold cross validation and grid search are applied to optimize the models and evaluate their performance.

4. Application to the decay of ^{11}Be

We have applied these new methods to the study of the decay of ^{11}Be, the most promising candidate to observe the rare β^--delayed proton emis-

sion. The β-delayed charged-particle emission of ^{11}Be is dominated by the α branch [13] and the possible proton branch is expected to be several orders of magnitude weaker. Evidence of such a proton branch was found in Ref. [14], recently supported by the determination of a near-threshold proton resonance in ^{11}B [15, 16].

The OTPC group performed an experiment at ISOLDE-CERN looking for this exotic decay branch. The gas mixture was chosen to measure the full α spectrum needed for normalization of the possible proton branch. In the analysis, it was found that, in such conditions, the possible low-energy protons are hardly distinguishable from low-energy α particles by energy loss fits [17], as illustrated in Fig. 2 (c). In order to shed light on this problem, we have trained ML models based on MC uniform energy distributions of α-^{7}Li and proton-^{10}Be in the range of 0–600 keV. A confusion matrix of the validation of these models is shown in Fig. 2 (a), showing an accuracy rate close to 90% to classify α and proton events. We have further investigated the performance of the models by applying them to a simulated data set of events sampled from the experimental α distribution and from the expected proton distribution. An accuracy rate above 80% is obtained, as shown in Fig. 2 (b). These accuracy rates together with the probability of belonging to each of the two classes given by the models will be used to determine a branching ratio for proton emission from experimental data.

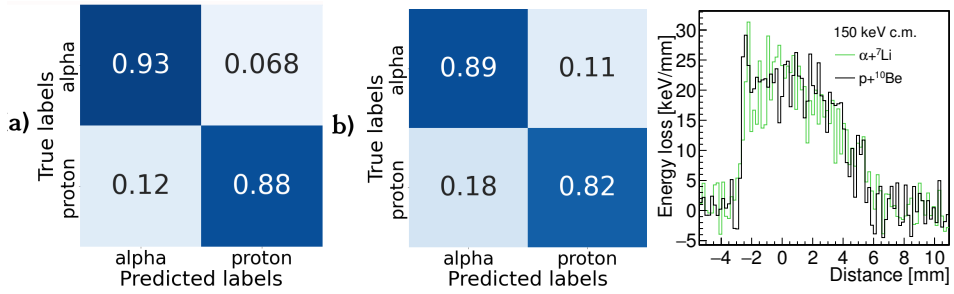

Fig. 2. Confusion matrices of the performance of the XGBoost model for the validation set after training on uniform energy distributions [panel (a)] and for a data set with realistic energy distributions [panel (b)]. In panel (c), we show MC simulations of the full track projection of proton and α emission for the OTPC experiment of ^{11}Be, considering the corresponding recoils and center-of-mass energy of 150 keV.

Finally, as the next steps, we are working in the inclusion of the β^{-}-delayed tritium channel, also energetically open in the decay of ^{11}Be, and we plan to optimize the models to enhance the accuracy rate at low energies.

This work has been supported by the Polish National Agency for Academic Exchange (NAWA) under grant No. PPN/ULM/2019/1/00220 and by the National Science Center, Poland (NCN) under contract No. 2019/35/D/ST2/02081.

REFERENCES

[1] M. Pfützner *et al.*, *Eur. Phys. J. A* **14**, 279 (2002).
[2] J. Giovinazzo *et al.*, *Phys. Rev. Lett.* **89**, 102501 (2002).
[3] M. Ćwiok *et al.*, *IEEE Trans. Nucl. Sci.* **52**, 2895 (2005).
[4] K. Miernik *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **581**, 194 (2007).
[5] M. Pomorski *et al.*, *Phys. Rev. C* **83**, 014306 (2011).
[6] M.P. Kuchera *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **940**, 156 (2019).
[7] R. Solli *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **1010**, 165461 (2021).
[8] S. Agostinelli *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **506**, 250 (2003).
[9] C.A. Douma *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **990**, 164951 (2021).
[10] P. Žugec *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **1033**, 166686 (2022).
[11] A. Allisy *et al.*, ICRU Report 49, International Commission of Radiation Units and Measurements, 1993.
[12] F. Chollet *et al.*, Keras, `https://keras.io`
[13] J. Refsgaard *et al.*, *Phys. Rev. C* **99**, 044316 (2019).
[14] Y. Ayyad *et al.*, *Phys. Rev. Lett.* **123**, 082501 (2019).
[15] Y. Ayyad *et al.*, *Phys. Rev. Lett.* **129**, 012501 (2022).
[16] E. Lopez-Saavedra *et al.*, *Phys. Rev. Lett.* **129**, 012502 (2022).
[17] N. Sokołowska, Ph.D. Thesis, University of Warsaw, 2023.