Capturing Neutrinos and New Physics at ProtoDUNE with a Machine Learning-Driven Trigger Algorithm

Ciaran Hasnip
CERN
ciaran.mark.hasnip@cern.ch

Abstract

The Deep Underground Neutrino Experiment (DUNE) will deploy massive liquid argon time projection chamber (LArTPC) detectors to study the properties of neutrinos with unprecedented precision. Currently, two full-scale prototypes of the DUNE far detectors (FDs) are at CERN, which are called the ProtoDUNE detectors. Recent studies indicate that these prototypes could have the potential to detect neutrinos and beyond the standard model (BSM) particles from one of the targets in CERN's north area that is exposed to the 400 GeV Super Proton Synchrotron (SPS) beam. A limiting factor in any neutrino and BSM physics program at ProtoDUNE is the ability to trigger for the interesting signal whilst rejecting the overwhelming cosmic-ray background present at these surface detectors. A new exclusive online self-triggering algorithm has been developed for the DUNE data acquisition (DAQ) system that uses an XGBoost gradient boosted decision tree (GBDT) model to select data with neutrino-like properties and reject cosmic-ray background. Simulations demonstrate that this algorithm improves the efficiency of selecting neutrino events at ProtoDUNE compared to existing algorithms. This algorithm has also been deployed in real data-taking operations at ProtoDUNE.

1 Neutrinos and new physics at ProtoDUNE

The Deep Underground Neutrino Experiment (DUNE) will measure the properties of neutrinos with unprecedented precision using liquid argon time projection chamber (LArTPC) detector technology [1]. The massive DUNE LArTPCs are currently under construction. Two full-scale prototypes that test and validate the technology of the future DUNE detectors operate at CERN, which are called the ProtoDUNE detectors. The two versions of ProtoDUNE are called ProtoDUNE-HD and ProtoDUNE-VD. [2]. LArTPCs produce high-resolution 3D images of particle interactions that allow for precise calorimetry and particle ID across a broad energy range.

There is an opportunity at the ProtoDUNE detectors to search for neutrinos and new beyond the standard model (BSM) signatures [3]. This is due to the alignment of the 400 GeV protons from the CERN Super Proton Synchrotron (SPS) beam impinging on one of the fixed targets in CERN's north area and the ProtoDUNE detectors downstream. This experimental setup is illustrated in Figure 1. The collision of protons on the fixed target produces a large flux of hadrons travelling towards the ProtoDUNE detectors. These hadrons will decay in flight to neutrinos or, possibly, BSM particles, such as a heavy neutral lepton (HNL). Such particles could interact or decay within the ProtoDUNE detectors, enabling the study of neutrinos and a search for new physics with the SPS beam. The expected neutrino energy spectrum is expected to peak at 5-10 Gev and extend out to over 100 GeV.

Submitted to 7th Fast Machine Learning for Science Conference (FastML 2026). Do not distribute.

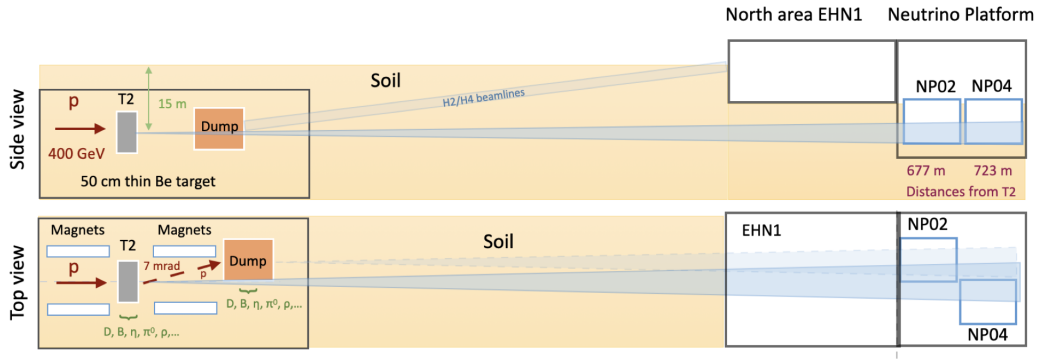


Figure 1: Illustration of the experimental setup. Protons from the SPS beam hit the "T2" target in CERN's north area. The resulting flux of neutrinos or hypothetical BSM particles is directed towards the ProtoDUNE-HD and ProtoDUNE-VD detectors, labelled here as NP04 and NP02 respectively. ProtoDUNE-HD and ProtoDUNE-VD are 723 m and 677 m downstream of the target respectively. Magnets surround the target facility, altering the direction of hadrons. This Figure is taken from [3].

2 DUNE data acquisition system

The detailed particle interaction data collected by LArTPCs results in large data volumes. Trigger algorithms that select data during real-time detector operation are therefore relied upon to significantly reduce data volumes sent to storage for offline analysis. The DUNE data acquisition system (DAQ) processes incoming LArTPC data through a hierarchical series of algorithms that finds clusters of charge on each detector channel, collects those clusters into regions of activity and creates trigger candidates that instructs the DAQ whether to record a piece of data or not. DUNE's real-time data-processing for the trigger has been demonstrated and validated at ProtoDUNE [4].

Capturing neutrinos and BSM physics at ProtoDUNE presents a particular challenge for triggering. Unlike the future underground detectors of DUNE, the ProtoDUNE detectors are at the surface, so are exposed to a significant flux of cosmic-rays. The trigger algorithms must therefore efficiently select rare neutrino and BSM signals whilst rejecting the overwhelming cosmic-ray background in order to control the volume of data sent to storage.

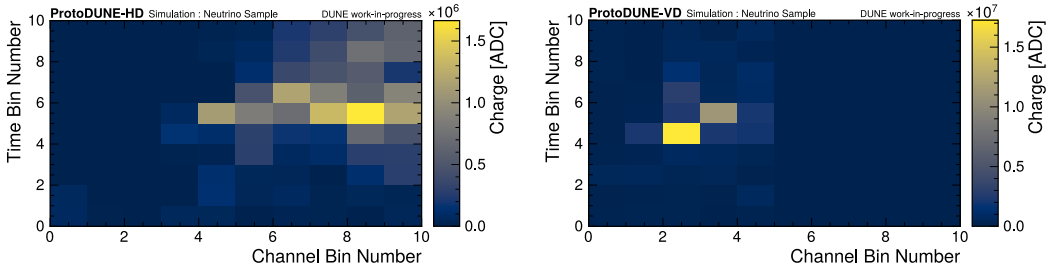
3 Machine learning trigger algorithm for ProtoDUNE

A machine learning (ML)-driven trigger algorithm has been developed for a rare event search at ProtoDUNE. The first step in the hierarchical series of trigger algorithms is to generate trigger primitive (TP) objects, which represent clusters of charge on a single detector channel. TPs are continuously generated and stream into the DAQ. An earlier version of the following algorithm was reported as a contribution to the ML-HEQUPP white paper [5]:

1. Calculate sum of the charge of the TPs in a 0.32 ms window of time. If it crosses a minimum threshold of 200,000 ADC units continue, else find the next TP.
2. Generate a coarse pixelated time versus channel number image of the TPs in the 0.32 ms window, where the amplitude of each pixel is the total charge of the TPs in the pixel.
3. Pass the image to an XGBoost ML model, which calculates the probability that the image is a neutrino. If the probability crosses a threshold, trigger ProtoDUNE.

This algorithm therefore uses the spatial and calorimetric information of LArTPCs to generate images of the activity in the detector from TPs, which are classified as either neutrino or cosmic-like in real-time using an XGBoost model [6]. Two independent XGBoost models were trained, one with simulated ProtoDUNE-HD data and the other with simulated ProtoDUNE-VD data. This allows the trigger algorithm to be deployed on either of the ProtoDUNE detectors. LArSoft was used to simulate particle interactions and the detector response [7]. Cosmic-ray background events were generated with CORSIKA [8]. The signal sample included a neutrino interaction simulated by GENIE in addition to the cosmic rays [9]. The simulated TPs were extracted from these events and binned as a

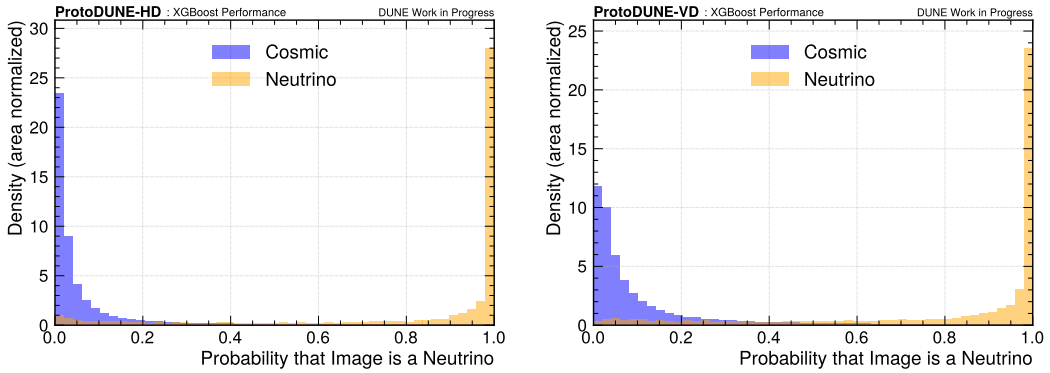
function of time and channel number to create an image where each time and channel bin is a pixel. The total time window for each binned sample was 0.32 ms. Approximately 50,000 neutrino samples and 900,000 cosmic-ray samples were produced each for ProtoDUNE-HD and ProtoDUNE-VD. Figure 2 shows examples of the signal training images.



(a) (b)

Figure 2: Figures 2a and 2b show examples of neutrino images used to train XGBoost models for ProtoDUNE-HD and ProtoDUNE-VD respectively. The colour scale of each pixel represents the sum of the charge from all the TPs in that pixel.

The parameters of the XGBoost models for both ProtoDUNE detectors were chosen to be 250 trees, with a max depth of 6 and a learning rate of 0.3. The parameters were chosen to balance model performance and complexity. A simpler model allows the XGBoost inference speed to meet the low-latency requirements of real-time data processing at ProtoDUNE. There are 101 features in both XGBoost models: one feature for each of the pixels in the ten-by-ten images shown in Figure 2 and a final feature for the total charge across the whole image.



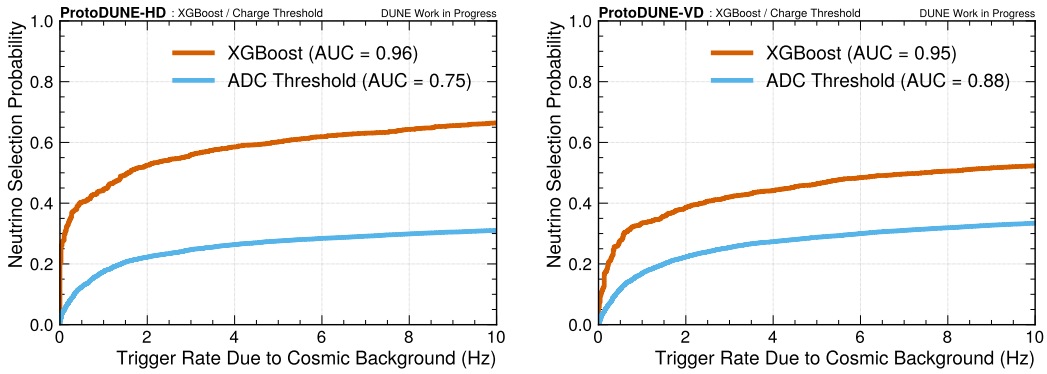
(a) (b)

Figure 3: Figures 3a and 3b show the predicted probability that the test sample images are a neutrino event for the ProtoDUNE-HD and ProtoDUNE-VD XGBoost models respectively.

3.1 Trigger algorithm performance

Of the neutrino and cosmic samples, 80% were used for training the XGBoost models and the remainder reserved for testing. The training data was filtered to remove images with charge less than 200,000 ADC. Figure 3 shows the relative performance of the trained ProtoDUNE-HD and ProtodUNE-VD XGBoost models in classifying images from the test samples as either cosmic or neutrino-like. Clear separation between signal and background can be seen in the case of both models.

The probability of the trigger selecting a neutrino as a function of the trigger rate due to cosmic-ray background for each XGBoost model is shown in Figure 4. The trigger rate is determined by how often cosmic background is selected. The best trigger algorithm previously available for a rare event search at ProtoDUNE relied on the sum of charge in a 0.32 ms time window. Figure 4 demonstrates significant improvements in neutrino selection probability when using the XGBoost models compared to the sum of the charge across all possible trigger rates.

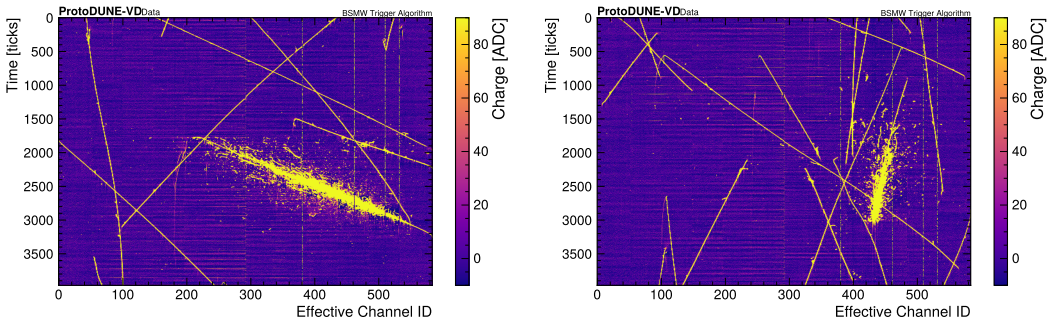


(a) (b)

Figure 4: Figures 4a and 4b show the neutrino selection probability as a function of the expected trigger rate at ProtoDUNE-HD and ProtoDUNE-VD respectively. In both detectors, significant improvement in physics performance are seen when using XGBoost compared to the charge threshold.

3.2 Deployment at ProtoDUNE

The XGBoost models have been integrated into the DUNE DAQ software, allowing for deployment of the trigger algorithm described in this paper at ProtoDUNE-VD for real data-taking operations. The trained XGBoost models were exported as compilable C++ code and built directly into the DUNE DAQ software, allowing for model inference times at microsecond-level that meet the low-latency requirements for triggering at ProtoDUNE. The trigger algorithm has successfully been operated and used to take data at ProtoDUNE-VD. Figure 5 shows two example event displays of some of the data recorded by this ML-driven trigger algorithm.



(a) (b)

Figure 5: Figures 5a and 5a show example event displays from a real data-taking run at ProtoDUNE-VD that used the ML-driven trigger algorithm described in this paper. The events are cosmic-ray background, but display features similar to that expected from signal.

4 Conclusions

A new ML-dirven trigger algorithm has been developed for the ProtoDUNE detectors at CERN, targeting neutrinos and BSM signatures. XGBoost models were trained to separate coarse images of neutrino and cosmic-ray interactions. Figure 4 indicates that this approach yields a significant improvement in performance over the previous non-ML trigger in both ProtoDUNE detectors. The ProtoDUNE-HD XGBoost model appears to perform better than the ProtoDUNE-VD case and this is due to the rotation of the ProtoDUNE-VD detector relative to the incoming neutrino flux. This study has demonstrated the feasibility of both deploying low-latency ML-driven trigger algorithms within the DUNE DAQ and of significantly improving the performance of the ProtoDUNE trigger for any future rare event search at the ProtoDUNE detectors.

References

[1] DUNE Collaboration, Babak Abi, et al. Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: DUNE Physics. 2 2020.

[2] Luis Manzanillas. Status of ProtoDUNE-II. In *Proceedings of 42nd International Conference on High Energy Physics — PoS(ICHEP2024)*, volume 476, page 219, 2024. doi: 10.22323/1.476.0219.

[3] Pilar Coloma, Jacobo López-Pavón, Laura Molina-Bueno, and Salvador Urrea. New physics searches using protodune and the cern sps accelerator. *Journal of High Energy Physics*, 2024(1): 1–18, 2024.

[4] DUNE Collaboration, Babak Abi, et al. The Single-Phase ProtoDUNE Technical Design Report. 6 2017.

[5] Julia Gonski et al. Machine Learning on Heterogeneous, Edge, and Quantum Hardware for Particle Physics (ML-HEQUPP). 2 2026.

[6] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794. ACM, August 2016. doi: 10.1145/2939672.2939785. URL `http://dx.doi.org/10.1145/2939672.2939785`.

[7] E.L. Snider and G. Petrillo. Larsoft: toolkit for simulation, reconstruction and analysis of liquid argon tpc neutrino detectors. *Journal of Physics: Conference Series*, 898(4):042057, 10 2017. doi: 10.1088/1742-6596/898/4/042057. URL `https://dx.doi.org/10.1088/1742-6596/898/4/042057`.

[8] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw. *CORSIKA: a Monte Carlo code to simulate extensive air showers.* 1998.

[9] C. Andreopoulos et al. The GENIE Neutrino Monte Carlo Generator. *Nucl. Instrum. Meth. A*, 614:87–104, 2010. doi: 10.1016/j.nima.2009.12.009.