# Catalog of Benchmarks, Software, and Datasets for the DUNE Project's Fast Machine Learning Initiatives from Gemini

## 1. Introduction

The Deep Underground Neutrino Experiment (DUNE) stands as a leading international endeavor in the field of neutrino science, aiming to unravel fundamental questions about neutrino oscillations, the potential for proton decay, and the characteristics of supernova neutrinos. The experiment's design incorporates massive liquid argon time projection chamber (LArTPC) detectors, which are expected to generate an unprecedented volume of data requiring advanced computational methods for effective analysis [1]. In this context, machine learning (ML) has become an increasingly vital tool in high-energy physics, offering solutions for a wide array of tasks, including the crucial areas of real-time event triggering, precise particle identification, accurate track reconstruction, and efficient reduction of background noise [3].

This report provides a systematic catalog of the benchmarks, software artifacts, and datasets that are associated with the Fast Machine Learning (FastML) efforts within the DUNE project, based on the provided research materials. The primary objective is to create a well-structured and detailed resource that outlines the key ML-related developments, available resources, and their measured performance metrics. This compilation aims to enhance understanding and facilitate collaboration within the DUNE research community and among interested parties in the broader scientific landscape. The information presented herein is exclusively derived from the provided URLs and their content, which includes presentation slides and any linked resources referenced within those materials. The methodology employed involves a thorough extraction of all mentions of benchmarks, software, and datasets, along with their corresponding descriptions, application areas, and pertinent technical specifications. This extracted information is then organized into a tabular format to address the user's request for an Excel spreadsheet representation.

## 2. Detailed Catalog of DUNE Artifacts

This section presents a comprehensive catalog of the benchmarks, software, and datasets identified within the research materials. The information is structured in a table to offer a clear and organized overview, directly addressing the user's requirement for an Excel spreadsheet format. Each row in the table represents a distinct artifact, with the columns detailing the specific attributes requested in the user query.

**DUNE FastML Artifact Catalog**

| Artifact Name | Description | Application Field | Nature of Model | Is model Deep Learning or not? | Size of model | Nature of Data | Modality of Data | Size of Data with units | Nature of benchmarks with machines used | Links to Papers | Links to data and software (GitHub) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| NVIDIA Triton | An online inference framework. | Fast AI/ML inference in DUNE Data Acquisition. | Used to query a server containing pre-loaded models. | Yes (serves deep learning models) | Not detailed. | Can process data in a format read by DUNE DAQ or reformatted data. | Not detailed. | Not detailed. | Testing inference rate with a "simple" model (adds two vectors). Observed increasing inferen | None mentioned. | None mentioned. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | ce time with more requests. Machines used not specified. | | |
| NuSonic framework | A framework implemented by Mike Wang et al. | Worked in LArSoft for DUNE. | Not detailed. | Not detailed. | Not detailed. | Reads data in a format used by LArSoft. | Not detailed. | Not detailed. | Not detailed. | None mentioned in this document. | Mentioned as working in the larrecodnn repository. |
| DUNE DAQ | A collection of cmake packages ma | Data acquisition for the Deep Undergr | Aims to integrate online inference usin | Not applicable. | Not applicable. | Reads raw data as "frames" (64 time tick | Raw detector signals. | Not detailed. | Currently testing Triton with fake data a | None mentioned. | None mentioned in this document. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | naged by Spack, an independent data acquisition system for DUNE. | ound Neutrino Experiment. | g models trained on its data format. | | | s x 64 channels), processed into TPs, "activities," and "candidates." TPs and Trigger Records written to HDF5. Can use configurable "fake data" trigger with TP generation | | | (HSI with 3 readout elements ~12 kHz of TPs), sending requests to a "simple" model every 100 TPs (~O(100) inference requests per second, each taking ~3 | | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | for testing. | | | ms). Machines used not fully specified (likely standard server hardware for Triton). | | |
| DUNE Near Detector (ND) data | Data from the near detector at Fermilab. | Long-baseline neutrino oscillation experiment. | Models will eventually be trained on this data. | Not applicable. | Not applicable. | Not fully detailed. | Not detailed. | Not detailed. | Data rates in Iceberg (DUNE cold electronics test stand at Fermilab) are a | None mentioned. | None mentioned. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | subj ect of futu re wor k. | | |
| DU NE Far Det ecto r (FD ) dat a | Dat a fro m the far det ecto r at SU RF. | Lon g-b asel ine neu trin o osci llati on exp eri me nt. | Mo dels will eve ntu ally be trai ned on this dat a. | Not appl icab le. | Not appl icab le. | Liqu id arg on time proj ecti on cha mb er (LAr TP C) tech nolo gy. | Det ecto r sign als. | Not det aile d. | Not det aile d. | Non e me ntio ned . | Non e me ntio ned . |
| Prot oD UN E-S P dat a | Dat a fro m the Prot oD UN E-S P det ecto r at CE RN. | Test ing LAr TP C tech nolo gy for DU NE. | Mo dels can be trai ned on this dat a. | Not appl icab le. | Not appl icab le. | Liqu id arg on time proj ecti on cha mb er (LAr TP C) tech nolo gy. | Det ecto r sign als. | Ra w trig ger rec ord s wer e aro und 75 MB com pre sse d. | Not det aile d. | Non e me ntio ned . | Non e me ntio ned . |
| 2D | A | Eve | Con | Yes. | 19, | Initi | Ima | Not | Perf | Non | Non |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CNN for radiological background rejection | 2D Convolutional Neural Network. | nt level trigger to reject frames containing only radiological background on raw data. | volutional Neural Network. | | 750 parameters. | ally simulated in LArSoft for raw data readout, now developing datasets and training techniques to perform inference on TP based images. | ges derived from detector signals. | detailed. | ormance on simulated data: 99.42% signal efficiency, 99.69% background rejection. GPU inference time for optimal batch: 80 us per image. FPGA inferen | e mentioned. | e mentioned. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | ce time : 850 us per eve nt. Res our ce usa ge on FP GA: LUT : 5.2 2%, BR AM: 18. 95 %, Reg ister : 4.2 6%, DS P: 8.1 7*. Mac hine s use d: GP U and FP GA. | | |
| Neu | A | Ide | Loc | No | Not | Not | Not | Not | AU | Non | Non |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| trino Interaction Identification (LBP+Dense) | model using Local Binary Patterns (LBP) and a dense layer. | ntifying neutrino interactions for supernova pointing. | al Binary Patterns (LBP) and a dense neural network. | (LBP is feature engineering). | detailed. | yet implemented in LArSoft. | detailed. | detailed. | C: 0.85, 83% eES->eES. GPU and FPGA inference times not measured or implemented in this context. | e mentioned. | e mentioned. |
| 1D CNN for finding Region of Interest (ROI) | A 1D Convolutional Neural Network. | Finding regions of interest in the data. | Convolutional Neural Network. | Yes. | Not detailed. | The ROIs look similar to "trigger primitives" (TP | 1D signal data. | Not detailed. | Not detailed. | Mentioned in the context of a presentation by Mike | None mentioned. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | s). | | | | Wang. | |
| 1D Autoencoder for denoising | A 1D Autoencoder. | Denoising data. | Autoencoder. | Yes. | Not detailed. | Used for electronics noise reduction. | 1D signal data. | Not detailed. | ~69% primary track efficiency, ~99% background rejection. GPU inference time: 0.061 sec/nuCC event, 0.031 sec/eES event. Machine used: | Mentioned in the context of a presentation by Mike Wang and J. Mitrevski's Fast ML talk from 2023. | None mentioned. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | GPU. | | |
| Spack | A package manager used to manage DUNE DAQ. | Management of software dependencies for DUNE DAQ. | Not applicable. | Not applicable. | Not applicable. | Not applicable. | Not applicable. | Not applicable. | Not applicable. | None mentioned. | None mentioned. |
| LArSoft | Software used for simulation and reconstruction in liquid argon TPC experime | Offline model development and simulation for DUNE. | Used for initial development and simulation of the 2D CNN. | Not applicable. | Not applicable. | Reads raw data in its own format. | Detector signals. | Not detailed. | Initial simulations of the 2D CNN were performed in LArSoft. | None mentioned. | None mentioned in this document. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | nts. | | | | | | | | | | |
| CNN for neutrino interaction classification | A Convolutional Neural Network. | Neutrino interaction classification in the DUNE far detector. | Convolutional Neural Network. | Yes. | Not fully detailed. | Data from the DUNE far detector, which utilizes liquid argon time projection chamber (LArTPC) technology. Supernova neutrino interactions appear as | Detector signals. | Not detailed. | Capable of exceeding its design sensitivity on simulated data. Machines used not specified. | Abi et al. (2020a), "Neutrino interaction classification with a convolutional neural network in the DUNE far detector"; Abi (2020), "Deep Underground Neu | Not mentioned in this snippet. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | small (tens of cm spatial scale). | | | | trino Experiment (DUNE) Far Detector Technical Design Report, Volume II DUNE Physics". | |
| Machine learning in DUNE's triggering framework | Machine learning techniques. | Triggering system for DUNE. | Not specified. | Potentially Yes. | Not detailed. | Data from the liquid argon time projection chamber. | Detector signals. | Not detailed. | Not detailed. | Ongoing efforts related to real-time event selection and reconst | Not mentioned in this snippet. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | ruction in neutrino detectors, including DUNE, with associated papers listed in the references of the article. | |
| GAr Soft | An art-based software suite used to simulate the | Simulation and reconstruction of events in gaseous | Includes simulation of ionization and scintillation, elec | No. | Not detailed. | Can import energy deposits from edep-sim. | Simulated particle interactions. | Not detailed. | Not detailed. | Not explicitly mentioned in the text. | Not explicitly mentioned in the text. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | ND-GAr detector. | argon time-projection chamber (GArTPC) detectors. | tron drift with diffusion and attenuation, and models of anode planes (IROCs and OROCs). | | | | | | | | |
| larnd-sim | ND-LAr detector simulation software written entirely in Python and compile | Simulation of the ND-LAr (pixel readout LArTPC) detector. | Takes into account quenching and drifting effects for ionization electrons in | No. | Not detailed. | Takes energy deposits from edep-sim as input (converted to HDF5 | Simulated particle interactions. | Not detailed. | Not detailed. | Not explicitly mentioned in the text. | Not explicitly mentioned in the text. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | d on the GPU using Numba for CUDA. | | liquid argon and calculates the current induced on pixels. Simulates electronics response and photon detection using a GPU-optimized lookup table. | | | format) and saves output in HDF5 format. | | | | | |
| edep-sim | A Geant4-bas | Simulating part | Based on Gea | No. | Not detailed. | Outputs ROOT | Simulated part | Not detailed. | Not detailed. | Not explicitly | Not explicitly |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | ed model used to propagate particles through the near detectors. | icle propagation and energy deposition in the near detector components. | nt4. | | | trees with energy deposit information and associated truth labels. | icle interactions. | | | mentioned in the text. | mentioned in the text. |
| Wire-Cell Toolkit (WCT) | Used for detailed 2D modeling of wire responses in wire-based LArTPCs. | Signal processing and reconstruction in wire-based LArTPCs. | Includes effects of induced currents on neighboring wires and folds in the electronics | No. | Not detailed. | Takes signal waveforms as input and produces a 3D model of ionization distributi | Detector signals. | Not detailed. | Not detailed. | References papers which likely contain more details. | Not explicitly mentioned in the text. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | response function. The wire-cell algorithm performs tomographic inversion of signals. | | | on. | | | | | |
| Pandora | A multi-algorithm approach to particle reconstruction. | Pattern recognition and event reconstruction, identifying clusters, tracks, and sho | Uses various algorithms to associate objects in 2D views to reconstruct 3D | No. | Not detailed. | Takes hits as input and outputs clusters, tracks, and showers stored in RO | Processed detector signals. | Not detailed. | Not detailed. | Reference paper likely contains more details. | Not explicitly mentioned in the text. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | wers in 3D. | objects. | | | OT trees. | | | | | |
| CAFAna (Common Analysis Format Analyzer) | A library of analysis tools developed by the NOvA collaboration. | Final stage physics analysis. | Works with simplified ROOT-based ntuples (CAF files). | No. | Not detailed. | Analyzes CAF files, which are reduced data samples. | Reduced physics data. | Not detailed. | Not detailed. | Reference papers likely contain more details. | Not explicitly mentioned in the text. |
| PPFX (Package to Predict the FluX) | A framework developed by the MINERvA collaboration used to estimate | Estimating neutrino flux uncertainties. | Reweighting framework for beamline simulations. | No. | Not detailed. | Used to reweight flux files. | Simulation parameters. | Not detailed. | Not detailed. | Reference papers likely contain more details. | Not explicitly mentioned in the text. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | uncertainties on the neutrino flux at the near detector. | | | | | | | | | | |
| DUNE ND GGD | A package that takes Python classes called builders as input and produces GDML files for the near | Creating geometry descriptions for the near detectors. | Uses Python classes to define detector geometry. | No. | Not detailed. | Outputs GDML files. | Detector geometry descriptions. | Not detailed. | Not detailed. | None mentioned in this snippet. | https://github.com/dune/dunendggd |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | detectors. | | | | | | | | | | |
| Geant4 Reweight | A reweighting framework developed by DUNE. | Adjusting existing Geant4 simulations based on new interaction cross-section knowledge. | Reweights Geant4 simulation output. | No. | Not detailed. | Operates on the output of Geant4 simulations. | Simulated particle interactions. | Not detailed. | Not detailed. | Reference paper likely contains more details. | Not explicitly mentioned in this snippet. |
| OverlayGenie | A package used for overlaying simulated interaction | Simulating backgrounds by combining simulated events | Combines simulated hits with raw data. | No. | Not detailed. | Requires libraries of properly-formatted raw data and sim | Detector signals and simulation data. | Not detailed. | Not detailed. | Not explicitly mentioned in this snippet. | Not explicitly mentioned in this snippet. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | information on real data. | with real detector data. | | | | ulated hits. | | | | | |
| ProtoDUNE-DP Data | Data collected from the dual-phase ProtoDUNE detector at CERN. | Testing the full data cataloging and reconstruction chain. | Not applicable. | Not applicable. | Not applicable. | Raw data files (run sequence files). | Detector signals. | 3.2 GB, each containing 30 trigger records. | Not detailed. | Not mentioned in this snippet. | Not mentioned in this snippet. |
| Public LArTPC Simulation Dataset | A publicly available dataset. | Evaluating machine learning-based reconstruction algorithms for | Not applicable. | Not applicable. | Not applicable. | A LArTPC simulation dataset. | Simulated particle interactions. | Not detailed. | Not detailed. | Reference paper likely contains details about this dataset | Not mentioned in this snippet. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ND-LAr. | | | | | | | | . | |

# 3. Summary and Initial Observations

The analysis of the provided materials reveals a significant effort to integrate machine learning methodologies into various aspects of the DUNE project [2]. This includes the development and application of deep learning models, such as Convolutional Neural Networks (CNNs) and Autoencoders, alongside more traditional machine learning techniques like those employing Local Binary Patterns (LBP) [4]. These efforts span a wide range of applications, from real-time data processing within the data acquisition system to offline analysis and simulation frameworks.

A prominent theme is the crucial role of simulation software in the DUNE ecosystem [2]. Frameworks like LArSoft, GArSoft, and larnd-sim are fundamental for both detector design and the development and testing of analysis algorithms, including those based on machine learning. The availability of data from the ProtoDUNE single-phase (SP) and dual-phase (DP) prototypes is also of paramount importance [2]. These datasets provide real-world examples of detector responses and are invaluable for validating simulation tools and analysis techniques before the full DUNE detectors become operational. The fact that raw trigger records from ProtoDUNE-SP were around 75 MB compressed and run sequence files from ProtoDUNE-DP were 3.2 GB each gives a sense of the scale of data being handled [2].

The development and benchmarking of models like the 2D CNN for radiological background rejection, which has been tested on both GPUs and FPGAs, highlight the project's focus on achieving efficient real-time processing [4]. The reported performance metrics, such as the 99.42% signal efficiency and 99.69% background rejection on simulated data, along with the inference times on different hardware, demonstrate a rigorous approach to evaluating the suitability of these models for deployment in the DUNE environment. Furthermore, the adaptation of models to work with the specific data formats of the DUNE DAQ system, such as the retraining of the 2D CNN with sparser Trigger Primitives, underscores a commitment to practical implementation.

The DUNE collaboration also leverages tools and frameworks developed by other neutrino experiments, such as CAFAna from NOvA and PPFX from MINERvA [2]. This collaborative approach within the high-energy physics community allows DUNE to benefit from the existing expertise and resources in the field. The use of a package manager like Spack for managing DUNE DAQ dependencies further emphasizes the adoption of sound software engineering practices for this complex project [4].

The variety of machine learning models under investigation, each tailored for specific tasks like neutrino interaction identification, region of interest finding, and noise reduction, suggests a comprehensive and strategic approach to applying these techniques to the unique challenges of neutrino physics [3]. The performance benchmarks reported for models like the 1D Autoencoder for denoising (e.g., ~69% primary track efficiency, ~99% background rejection, and specific inference times) provide quantitative measures of their effectiveness.

# 4. Conclusion

This report has presented a detailed catalog of the benchmarks, software artifacts, and datasets associated with the Fast Machine Learning initiatives within the DUNE project, as documented in the provided research materials. These resources are essential for the ongoing development and preparation for the DUNE experiment. The benchmarks offer crucial performance metrics, the software artifacts enable vital simulation, reconstruction, and analysis capabilities, and the datasets serve as fundamental resources for training and validating the next generation of analysis algorithms.

The continued development and refinement of these machine learning-related tools and resources are vital for enabling DUNE to achieve its ambitious scientific objectives. The integration of advanced computational techniques, particularly machine learning, is poised to play a critical role in the success of this groundbreaking experiment, facilitating the extraction of meaningful insights from the vast and complex datasets that DUNE will generate. Further exploration of the presentations and papers referenced in the materials, as well as the investigation of repositories like larrecodnn, would likely provide even more granular details about the ongoing advancements in this dynamic field within the DUNE collaboration.

### Works cited

1. Exploring the Quantum Universe Pathways to Innovation and Discovery in Particle Physics, accessed March 15, 2025,
https://www.usparticlephysics.org/2023-p5-report/assets/pdf/2023_P5_Report_Spreads.pdf
2. indico.fnal.gov, accessed March 15, 2025,
https://indico.fnal.gov/event/56055/contributions/249571/attachments/159415/209586/Computing_CDR_5Sept2022.pdf
3. Applications and Techniques for Fast Machine Learning ... - Frontiers, accessed March 15, 2025, https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2022.787421/full
4. indico.fnal.gov, accessed March 15, 2025,
https://indico.fnal.gov/event/66520/contributions/301423/attachments/182439/250508/fast_ml_dunedaq_sonic_10_15_24.pdf