Machine Learning Techniques and Software for Neutrino Physics

C. Adams[1], A. Aurisano[2], J. Bian[3], M. Del Tutto[4], P. de Perio[5], N. Prouse[5], K. Terao*[6], and T. Wongjirad[7]

[1]Argonne National Laboratory
[2]University of Cincinnati
[3]University of California, Irvine
[4]Fermi National Accelerator Laboratory
[5]TRIUMF
[6]SLAC National Accelerator Laboratory
[7]Tufts University

Thematic Areas: (check all that apply □/■)
■ (CompF1) Experimental Algorithm Parallelization
□ (CompF2) Theoretical Calculations and Simulation
■ (CompF3) Machine Learning
■ (CompF4) Storage and processing resource access (Facility and Infrastructure R&D)
□ (CompF5) End user analysis
□ (CompF6) Quantum computing
□ (CompF7) Reinterpretation and long-term preservation of data and code
□ (Other) *[Please specify frontier/topical group]*

Machine learning (ML) techniques have been explored in most, if not all, aspects of neutrino physics research, in particular for experimental data reconstruction, pattern recognition, and analysis. For the most part, the R&D of machine learning applications, including both algorithms and software tools for distributed computing, have been developed within each experimental collaboration. Progress within the field of machine learning and neutrino physics has been impressive in a short time, yet each result is siloed into it's own experimental program and cross-cutting breakthroughs have not yet materialized.

Neutrino Physics and Machine Learning (NPML) workshops [1, 2] were called recently to gather a survey of research work in this area across the field of neutrino physics. As part of the workshop, the topics below emerged as common research thrusts across experimental programs that participated in the workshops.

- **Physics-informed machine learning** methods, or algorithms with inductive biases from domain experts, are actively being pursued. Examples include a detector geometry (e.g. hexagonal or spherical kernel for convolutional neural networks), conservation laws and causalities (e.g. tree structure for particle flow reconstruction).

- **Uncertainty estimation** including the ML model intrinsic uncertainty, methods to propagate uncertainty of input model parameters, and systematic uncertainty to account for physics mis-modeling, are common interests. Sharing expertise of existing techniques (e.g. Gaussian process, Bayesian neural networks, Monte Carlo drop-outs, etc.) and R&D beyond these approaches are both in critical need.

*Corresponding Author: Kazuhiro Terao (kterao@slac.stanford.edu)

- **Domain discrepancy mitigation**, in particular data and simulation, is considered critical since many existing ML applications for data reconstruction and analysis rely on the presence of high fidelity simulation tools for network training. Research progress is needed for both quantification of domain discrepancy and mitigation techniques.

- **Interpretable, end-to-end** data reconstruction and physics inference continue to be pursued. The interpretability is sought through uncertainty estimation and model architectures with a causal structure and inductive biases. Combining those models, including ML and non-ML algorithms, and optimizing the whole chain is a shared research challenge with high impact.

- **Utilization of high performance computing facilities** is of high interest especially with ML algorithms for which scalability through parallelism can be realized more so than traditional experimental software. This also has huge impact in accelerating ML algorithms development (e.g. *training* time reduction by orders of magnitude changed development workflow) and experimental data production.

- **Public dataset** for reproducibility and open, collaborative development is crucial. There is a critical value in sharing tools and design patterns for making a successful public dataset to foster development of an inter-experimental collaboration, and we strongly recommend to initiate a community-wide support.

- **Education** is considered lacking and impactful. It should be supported beyond workshops and summer schools, which are typically opt-in by individuals with limited travel funding support and only for a few weeks at the maximum duration.

We are currently preparing a white paper as a workshop product, which summarizes the survey of current R&D of ML applications and necessary resource for the future development and production using such applications (**CompF-01,03,04**). This white paper also includes a list of challenges in domain science that are good fit to be addressed using ML techniques to maximize the research impact (NF-01,02,03,06). with computing-specific natures to support the R&D effort of ML applications. In particular, as a product of this white paper and workshop we are building cross-cutting collaboration connections to enable all experiments using machine learning in neutrino physics to benefit.

References

[1] Neutrino 2020 satellite machine learning workshop. https://indico.slac.stanford.edu/event/377/, July 2020.

[2] Neutrino physics and machine learning. https://indico.slac.stanford.edu/event/371/, July 2020.