Snowmass2021 - Letter of Interest

The Future of Machine Learning in Rare Event Searches

Thematic Area: CompF3: Machine Learning
Additional Thematic Areas: CompF2: Theoretical Calculations and Simulation
CF1: Dark Matter, Particle-like; CF2: Dark Matter, Wave-like; NF1: Neutrino Oscillations
NF5: Neutrino properties; IF3: Solid State Detectors and Tracking; IF8: Noble Elements

Contact Information:
Scott Kravitz (Lawrence Berkeley National Laboratory) [swkravitz@lbl.gov]
Maria Elena Monzani (SLAC National Accelerator Laboratory) [monzani@stanford.edu]
Igor Ostrovskiy (University of Alabama) [iostrovskiy@ua.edu]

Authors: Complete list at the end of document.

Abstract: Rare event searches in dark matter and neutrino experiments share the unique challenge of identifying a few potential signal events from a much larger dataset. Future iterations of these experiments will need to collaborate to develop new techniques for addressing the challenges from increasingly-complex datasets. Machine learning (ML) methods are well-suited to many of these challenges, but require refinement in some cases and entirely new approaches in others to meet these needs. Opportunities for development include methods to increase reliability of ML techniques, through quantified uncertainties and human-interpretable results; unsupervised anomaly finding for unexpected or difficult to model backgrounds; as well as faster and more accurate simulations. Doing this well will require training of experts within this community, through domain-specific workshops and courses, in addition to state-of-the-art tools and hardware.

Introduction: Recent advances in Machine Learning (ML) have enabled sophisticated analyses of large scientific datasets. The HEP community has successfully leveraged these advances to improve many aspects of analysis including event reconstruction, particle identification, fast simulation, and triggering [1–3]. Similar approaches are gaining traction in the rare event search community and show promise in addressing some of the challenges facing us moving forward. Work across collaborations is vital to discern the field-specific advantages of ML, as well as to build confidence in and understanding of its results, so that it can be widely and effectively used.

Existing dark matter (DM) and neutrino experiments have demonstrated improvements in background rejection using simple neural networks and boosted decision trees [4, 5]. Deep learning techniques have also been successfully employed for reconstruction of physical quantities, such as energy and position [6], as well as particle identification [7, 8] and signal/background discrimination [9]. ML algorithms have also begun to make their way into hardware, through their implementation on FPGAs for triggering [10]. Despite these initial successes, using ML to its full potential will require addressing several challenges at the forefront of scientific computing.

Uncertainties and Interpretability: Core among these challenges is the need to establish ML-based approaches that can match traditional analyses in their reliability. This is especially relevant for algorithms utilizing low-level information, where building physical intuition about their inner workings is most difficult and the fidelity of simulations is challenging to verify. Techniques to reduce reliance on simulation [11] and demonstrate robustness against differences between simulation and data [12] are being explored, but must be further developed and tested. Data-driven training sets provide the ultimate in reliability. New techniques leveraging calibration data to generate such data sets would enable expansion of the domain of application of ML in our field.

Beyond improvements in training, quantitative establishment of uncertainties in ML outputs is integral to translating results to a final, rigorous statistical analysis. Additionally, tools to improve interpretability of ML outputs can allow human analyzers to benefit from their insights, re-examining their assumptions about what information is physically relevant. This emphasis on understanding both uncertainties and human-interpretable insights distinguishes physics from industrial applications; truly pioneering work is necessary to meet our goals.

Anomaly Finding and Events at Threshold: Rare event searches present a common computational challenge: identify a handful of possible "signal" events from several petabytes of background data. As these experiments grow in size and complexity, the potential for unexpected backgrounds increases. While anomaly detection methods are being explored in the accelerator physics context [13], their direct application to rare event searches is often limited due to their reliance on larger sample sizes. Exploration of unsupervised learning techniques for rare event searches is valuable here, as this class of algorithms does not require a priori knowledge of each event topology. Similarly, anomaly finding algorithms can augment development of traditional techniques by efficiently finding outliers in high-dimensional space, which may appear as a result of a bug or unexpected corner case in simulation or reconstruction.

Machine-learning techniques can also be effective for the reconstruction of very-low energy events and events close to the detector threshold, in which the available information is exceedingly sparse and statistical fluctuations are predominant. Improvements in reconstruction of low-energy events have been demonstrated in liquid xenon detectors [14, 15] and show promise for further study.

Simulations: Simulations are another area with strong potential for improvement from ML due to their intensive use of computing resources. In particular, as analyses delve into use of raw detector-level information, accurate simulation becomes both more challenging and more time-consuming. Generative models have shown potential in this realm [16], with significantly-reduced computing requirements in the generation stage, but remain relatively unexplored in the rare event community. An important open challenge for this approach will be extending generation to event types which are not well-modeled (if trained using simulation) or which are not well-represented in the underlying distribution (if trained using calibration data).

Tool Development and Training: More generally, gaining access to state-of-the-art software tools and hardware resources specific to deep learning, such as clusters of graphical and tensor processing units, will help ensure competitiveness with other related fields. As software and resources evolve, we must ensure our experimental data remains compatible with them. This may require community development of shared tools, following the example of software packages such as `uproot` [17]. Standardized datasets for rare event searches, analogous to the MNIST database of handwritten digits in computer vision [18], can also help with benchmarking of both algorithm performance and computing resources. The Noble Element Simulation Technique (NEST) software is designed for efficient simulation of noble element detectors and is used by multiple collaborations [19]; it could generate such shared datasets, enabling development of increasingly sophisticated ML models with the incorporation of low-level information.

Lastly, algorithms currently used in the field are largely based on techniques established in the private technology sector, while our unique problem space may lend itself to yet-unheard-of approaches. The development of new algorithms, ideally in collaboration with ML experts outside the field, has the potential to significantly improve performance on problems that do not readily map onto established paradigms.

To reach these goals, our field must begin training ML experts of our own. General ML resources are readily available online, but are not as efficient as field-specific training. To address this, we recommend dedicated ML workshops [20], where expertise can be shared through interactive tutorials, and promising techniques in the field can be discussed. These would also provide opportunities for collaboration with external experts. Development of physics-specific ML courses can also address this need.

[1] K. Albertsson *et al.*, "Machine learning in high energy physics community white paper," arXiv:1807.02876 [physics.comp-ph].

[2] A. J. Larkoski, I. Moult, and B. Nachman, "Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning," *Phys. Rept.* **841** (2020) 1–63, arXiv:1709.04464 [hep-ph].

[3] D. Guest, K. Cranmer, and D. Whiteson, "Deep Learning and its Application to LHC Physics," *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 161–181, arXiv:1806.11484 [hep-ex].

[4] **EXO-200 Collaboration** Collaboration, J. B. Albert *et al.*, "Search for $2\nu\beta\beta$ decay of $^{136}\mathrm{Xe}$ to the 0^+_1 excited state of $^{136}\mathrm{Ba}$ with the exo-200 liquid xenon detector," *Phys. Rev. C* **93** (2016) 035501.

[5] **SuperCDMS** Collaboration, R. Agnese *et al.*, "Search for Low-Mass Weakly Interacting Massive Particles with SuperCDMS," *Phys. Rev. Lett.* **112** no. 24, (2014) 241302, arXiv:1402.7137 [hep-ex].

[6] S. Delaquis, M. Jewell, I. Ostrovskiy, M. Weber, T. Ziegler, J. Dalmasson, L. Kaufman, T. Richards, *et al.*, "Deep neural networks for energy and position reconstruction in EXO-200," *Journal of Instrumentation* **13** no. 08, (2018) P08023–P08023.

[7] A. Aurisano, A. Radovic, D. Rocco, A. Himmel, M. D. Messier, E. Niner, G. Pawloski, F. Psihas, A. Sousa, and P. Vahle, "A Convolutional Neural Network Neutrino Event Classifier," *JINST* **11** no. 09, (2016) P09001, arXiv:1604.01444 [hep-ex].

[8] **MicroBooNE** Collaboration, R. Acciarri *et al.*, "Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber," *JINST* **12** no. 03, (2017) P03011, arXiv:1611.05531 [physics.ins-det].

[9] **EXO-200** Collaboration, G. Anton *et al.*, "Search for Neutrinoless Double-β Decay with the Complete EXO-200 Dataset," *Phys. Rev. Lett.* **123** no. 16, (2019) 161802.

[10] J. Duarte *et al.*, "FPGA-accelerated machine learning inference as a service for particle physics computing," *Comput. Softw. Big Sci.* **3** no. 1, (2019) 13, arXiv:1904.08986 [physics.data-an].

[11] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, "Learning to classify from impure samples with high-dimensional data," *Phys. Rev. D* **98** (2018) 011502.

[12] G. Perdue *et al.*, "Reducing model bias in a deep learning classifier using domain adversarial neural networks in the MINERvA experiment," *Journal of Instrumentation* **13** no. 11, (2018) P11020–P11020.

[13] J. H. Collins, K. Howe, and B. Nachman, "Anomaly Detection for Resonant New Physics with Machine Learning," *Phys. Rev. Lett.* **121** no. 24, (2018) 241803, arXiv:1805.02664 [hep-ph].

[14] S. Chan and LUX Collaboration, "Machine Learning Applied To Background Events Identification In LUX Dark Matter Experiment," in *APS April Meeting Abstracts*, vol. 2019 of *APS Meeting Abstracts*, p. T13.004. Jan., 2019.

[15] C. K. Khosa, L. Mars, J. Richards, and V. Sanz, "Convolutional Neural Networks for Direct Detection of Dark Matter," *J. Phys. G* **47** no. 9, (2020) 095201, arXiv:1911.09210 [hep-ph].

[16] A. Butter, T. Plehn, and R. Winterhalder, "How to GAN LHC Events," *SciPost Phys.* **7** no. 6, (2019) 075, arXiv:1907.03764 [hep-ph].

[17] J. Pivarski, P. Das, C. Burr, D. Smirnov, M. Feickert, T. Gal, L. Kreczko, N. Smith, N. Biederbeck, O. Shadura, M. Proffitt, benkrikler, H. Dembinski, H. Schreiner, J. Rembser, M. R., C. Gu, J. Rübenach, M. Peresano, and R. Turra, "scikit-hep/uproot: 3.12.0,". https://doi.org/10.5281/zenodo.3952728.

[18] Y. LeCun and C. Cortes, "MNIST handwritten digit database,". http://yann.lecun.com/exdb/mnist/.

[19] M. Szydagis, J. Balajthy, J. Brodsky, J. Cutter, J. Huang, E. Kozlova, B. Lenardo, A. Manalaysay, D. McKinsey, M. Mooney, J. Mueller, K. Ni, G. Rischbieter, M. Tripathi, C. Tunnell, V. Velan, and Z. Zhao, "Noble element simulation technique,". https://doi.org/10.5281/zenodo.3905382.

[20] *Dark-matter and Neutrino Computation Explored (DANCE) Machine Learning Workshop 2020*. 2020. https://indico.physics.lbl.gov/event/1192/.

Full Author List:

J. Aalbers,[1] D.S. Akerib,[2,3] A. Alqahtani,[4] S.K. Alsum,[5] C.S. Amarasinghe,[6] A. Ames,[2,3] P. An,[7] S.J. Andaloro,[8] T.J. Anderson,[2,3] N. Angelides,[9] H.M. Araújo,[10] J.E. Armstrong,[11] M. Arthurs,[6] X. Bai,[12] J. Balajthy,[13] S. Balashov,[14] J. Bang,[4] P.S. Barbeau,[15] J.W. Bargemann,[16] A. Baxter,[17] J. Bensinger,[18] E.P. Bernard,[19,20] A. Bernstein,[21] A. Bhatti,[11] A. Biekert,[19,20] T.P. Biesiadzinski,[2,3] R. Biondi,[22] H.J. Birch,[6] G.M. Blockinger,[23] K.E. Boast,[24] E. Bodnia,[16] B. Boxer,[17] P.A. Breur,[25] C.A.J. Brew,[14] P. Brás,[26] S. Burdin,[17] J.K. Busenitz,[27] M. Buuck,[2,3] R. Cabrita,[26] C. Carels,[24] D.L. Carlsmith,[5] M.C. Carmona-Benitez,[28] M. Cascella,[9] C. Chan,[4] N.I. Chott,[12] M. Clark,[29] A. Cole,[20] M.V. Converse,[30] A. Cottle,[24,31] G. Cox,[28] O. Creaner,[20] J.E. Cutter,[13] C.E. Dahl,[32,31] V. D'Andrea,[33] L. de Viveiros,[28] A. Depoian,[29] M.D. Diamond,[34] J.E.Y. Dobson,[9] E. Druszkiewicz,[30] T.K. Edberg,[11] S.R. Eriksen,[35] A. Fan,[2,3] A.D. Ferella,[33] S. Fiorucci,[20] H. Flaecher,[35] E.D. Fraser,[17] T. Fruth,[9] R.J. Gaitskell,[4] J. Genovesi,[12] C. Ghag,[9] E. Gibson,[24] M.G.D. Gilchriese,[20] S. Gokhale,[36] M.G.D.van der Grinten,[14] C.R. Hall,[11] C. Hardy,[2,3] S.J. Haselschwardt,[20] S.A. Hertel,[37] A. Higuera,[8] J.Y-K. Hor,[27] M. Horn,[38] D.Q. Huang,[4] C.M. Ignarra,[2,3] O. Jahangir,[9] W. Ji,[2,3] J. Johnson,[13] A.C. Kaboth,[39,14] A. Kamaha,[23] K. Kamdin,[20,19] K. Kazkaz,[21] D. Khaitan,[30] A. Khazov,[14] I. Khurana,[9] D. Kodroff,[28] A. Kopec,[29] L. Korley,[6] E.V. Korolkova,[40] H. Kraus,[24] S. Kravitz,[20] L. Kreczko,[35] B. Krikler,[35] V.A. Kudryavtsev,[40] R.F. Lang,[29] E.A. Leason,[41] K.T. Lesko,[20] C. Levy,[23] J. Li,[42] S. Li,[43] Z. Li,[44] J. Liao,[4] F.-T. Liao,[24] C. Liebenthal,[8] J. Lin,[19,20] A. Lindote,[26] R. Linehan,[2,3] W.H. Lippincott,[16,31] X. Liu,[41] C. Loniewski,[30] M.I. Lopes,[26] E. Lopez Asamar,[26] B. López Paredes,[10] W. Lorenzon,[6] S. Luitz,[2] P.A. Majewski,[14] A. Manalaysay,[20] L. Manenti,[9] R.L. Mannino,[5] N. Marangou,[10] M.F. Marzioni,[41] M.E. McCarthy,[30] D.N. McKinsey,[19,20] J. McLaughlin,[32] Y. Meng,[27] E.H. Miller,[2,3] E. Mizrachi,[11] A. Monte,[16,31] M.E. Monzani,[2,3] J.A. Morad,[13] J.D. Morales Mendoza,[2,3] E. Morrison,[12] B.J. Mount,[45] A.St.J. Murphy,[41] B.P. Nachman,[46] D. Naim,[13] A. Naylor,[40] C. Nedlik,[37] H.N. Nelson,[16] F. Neves,[26] J.A. Nikoleyczik,[5] A. Nilima,[41] I. Olcina,[19,20] K.C. Oliver-Mallory,[10] A. Oranday,[8] I. Ostrovskiy,[47] S. Pal,[26] K.J. Palladino,[5,24] J. Palmer,[39] N. Parveen,[23] S.J. Patton,[20] E.K. Pease,[20] B. Penning,[6] G. Pereira,[26] J. Pershing,[21] A. Piepke,[27] Y. Qie,[30] J. Qin,[29] J. Reichenbacher,[12] C.A. Rhyne,[4] Q. Riffard,[19,20] G.R.C. Rischbieter,[23] R. Rosero,[36] P. Rossiter,[40] L. Sanchez,[8] D. Santone,[39] A.B.M.R. Sazzad,[27] R.W. Schnee,[12] P.R. Scovell,[14] S. Shaw,[16] T.A. Shutt,[2,3] J.J. Silk,[11] C. Silva,[26] W. Skulski,[30] R. Smith,[19,20] M. Solmaz,[16] V.N. Solovov,[26] P. Sorensen,[20] J. Soria,[19,20] I. Stancu,[27] A. Stevens,[24] K. Stifter,[2,3] B. Suerfu,[19,20] T.J. Sumner,[10] N. Swanson,[4] M. Szydagis,[23] M. Tan,[24] W.C. Taylor,[4] R. Taylor,[10] D.J. Temples,[32] P.A. Terman,[48] D.R. Tiedt,[11] M. Timalsina,[12] W.H. To,[2,3] A. Tomás,[10] M. Tripathi,[13] D.R. Tronstad,[12] C.E. Tull,[20] C. Tunnell,[8] W. Turner,[17] L. Tvrznikova,[49,19] U. Utku,[9] A. Vaitkus,[4] V. Velan,[50] C.O. Vuosalo,[5] J.J. Wang,[6] B. Wang,[27] W. Wang,[37] J.R. Watson,[19,20] R.C. Webb,[48] R.G. White,[2,3] T.J. Whitis,[16,2] M. Williams,[6] M.S. Witherell,[20,19] F.L.H. Wolfs,[30] D. Woodward,[28] C.J. Wright,[35] X. Xiang,[4] J. Xu,[21] L. Yang,[51] M. Yeh,[36] and P. Zarzhitsky[27]

(Collaborations: LZ, XENON, DARWIN, EXO/nEXO, LUX, COHERENT, ATLAS)

[1]*Oskar Klein Centre, Department of Physics, Stockholm University, AlbaNova, Stockholm SE-1069*
[2]*SLAC National Accelerator Laboratory, Menlo Park, CA 94025-7015, USA*
[3]*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, Stanford, CA 94305-4085 USA*
[4]*Brown University, Department of Physics, Providence, RI 02912-9037, USA*
[5]*University of Wisconsin-Madison, Department of Physics, Madison, WI 53706-1390, USA*
[6]*University of Michigan, Randall Laboratory of Physics, Ann Arbor, MI 48109-1040, USA*
[7]*Department of Physics, Duke University, Science Dr, Durham, NC 27710, USA*

[8]Department of Physics & Astronomy, Rice University, 6100 Main St, Houston, TX 77005, USA
[9]University College London (UCL), Department of Physics and Astronomy, London WC1E 6BT, UK
[10]Imperial College London, Physics Department, Blackett Laboratory, London SW7 2AZ, UK
[11]University of Maryland, Department of Physics, College Park, MD 20742-4111, USA
[12]South Dakota School of Mines and Technology, Rapid City, SD 57701-3901, USA
[13]University of California, Davis, Department of Physics, Davis, CA 95616-5270, USA
[14]STFC Rutherford Appleton Laboratory (RAL), Didcot, OX11 0QX, UK
[15]Department of Physics, Duke University, and Triangle Universities Nuclear Laboratory, Science Dr, Durham, NC 27710, USA
[16]University of California, Santa Barbara, Department of Physics, Santa Barbara, CA 93106-9530, USA
[17]University of Liverpool, Department of Physics, Liverpool L69 7ZE, UK
[18]Brandeis University, Department of Physics, Waltham, MA 02453, USA
[19]University of California, Berkeley, Department of Physics, Berkeley, CA 94720-7300, USA
[20]Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA 94720-8099, USA
[21]Lawrence Livermore National Laboratory (LLNL), Livermore, CA 94550-9698, USA
[22]INFN Laboratori Nazionali del Gran Sasso, 67010 Assergi (AQ), Italy
[23]University at Albany (SUNY), Department of Physics, Albany, NY 12222-1000, USA
[24]University of Oxford, Department of Physics, Oxford OX1 3RH, UK
[25]SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA
[26]Laboratório de Instrumentação e Física Experimental de Partículas (LIP), University of Coimbra, P-3004 516 Coimbra, Portugal
[27]University of Alabama, Department of Physics & Astronomy, Tuscaloosa, AL 34587-0324, USA
[28]Pennsylvania State University, Department of Physics, University Park, PA 16802-6300, USA
[29]Department of Physics and Astronomy, Purdue University, West Lafayette, IN 47907, USA
[30]University of Rochester, Department of Physics and Astronomy, Rochester, NY 14627-0171, USA
[31]Fermi National Accelerator Laboratory (FNAL), Batavia, IL 60510-5011, USA
[32]Northwestern University, Department of Physics & Astronomy, Evanston, IL 60208-3112, USA
[33]University of L'Aquila, Via Vetoio, 67100 Coppito (AQ), Italy
[34]Department of Physics, University of Toronto, Toronto ON Canada M5S1A7
[35]University of Bristol, H.H. Wills Physics Laboratory, Bristol, BS8 1TL, UK
[36]Brookhaven National Laboratory (BNL), Upton, NY 11973-5000, USA
[37]University of Massachusetts, Department of Physics, Amherst, MA 01003-9337, USA
[38]South Dakota Science and Technology Authority (SDSTA), Sanford Underground Research Facility, Lead, SD 57754-1700, USA
[39]Royal Holloway, University of London, Department of Physics, Egham, TW20 0EX, UK
[40]University of Sheffield, Department of Physics and Astronomy, Sheffield S3 7RH, UK
[41]University of Edinburgh, SUPA, School of Physics and Astronomy, Edinburgh EH9 3FD, UK
[42]IBS Center for Underground Physics (CUP), Yuseong-gu, Daejeon, KOR
[43]Physics Department, University of Illinois, Urbana-Champaign, Illinois 61801
[44]Wright Laboratory, Department of Physics, Yale University, New Haven, Connecticut 06511, USA
[45]Black Hills State University, School of Natural Sciences, Spearfish, SD 57799-0002, USA
[46]Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
[47]Department of Physics and Astronomy, University of Alabama, Tuscaloosa, AL 35487, USA
[48]Texas A&M University, Department of Physics and Astronomy, College Station, TX 77843-4242, USA
[49]Yale University, Department of Physics, New Haven, CT 06511-8499, USA
[50]Department of Physics, University of California Berkeley, Berkeley, California 94720, USA
[51]Department of Physics, University of California San Diego, La Jolla, CA 92093