***Snowmass Letter of Interest*, Topic Areas: CompF03, CompF07**

# Machine Learning References for HEP

**Gabriele Benelli[1], Lindsey Gray[2], Gregor Kasieczka[3], Matthias Komm[4], Mia Liu[5], Scarlet Norberg[6], Alexx Perloff[7], Marcel Rieger[4], Savannah Thais[8], and Jean-Roch Vlimant[9]**

[1]Brown University, Providence, RI 02912, USA
[2]Fermi National Accelerator Laboratory, Batavia, IL 60510, USA
[3]Universität Hamburg, Institut für Experimentalphysik, 22761 Hamburg, Germany
[4]European Organization for Nuclear Research, Geneva, Switzerland
[5]Purdue University, West Lafayette, IN 47907, USA
[6]University of Puerto Rico Mayagüez, Mayagüez, Puerto Rico
[7]University of Colorado Boulder, Boulder, CO 80309, USA
[8]Princeton University, Princeton, New Jersey 08544, USA
[9]California Institute of Technology, Pasadena, California 91125, USA

## ABSTRACT

More and more, machine learning (ML) is being used within the high energy physics (HEP) domain. There are a vast number of ML applications and methodologies being used on a wide variety of topics. The growing number of ML applications within the HEP community naturally leads to an increased need to cite up-to-date and state-of-the-art ML papers, often found in conference proceedings or computer science journals. While the community has been successfully able to adapt ML techniques to our field, there isn't yet an easy way to reference the relevant ML papers for citation purposes. In this LoI, we will summarize the basic need for an up-to-date, cross referenced list of ML papers and some ways in which this goal can be achieved.

## 1 Introduction and Existing Resources

Fundamental particle physics experiments are known for state-of-the-art machines and 'big data' analyses which utilize high dimensional, complex datasets with the goal of discovering new features. To address these questions, HEP scientists, engineers, and technicians are frequently turning to machine learning (ML) techniques in the course of their research. These HEP-ML applications have been extensively documented in individual papers as well as in reviews, such as in [1–4]. These documents demonstrate the level of impact that ML techniques continue to have on accelerator and detector operations, data simulation, reconstruction, and analysis. While the use of ML has evolved rapidly within the HEP community and is oftentimes at the cutting edge of the field, it is an active area of research within the computer science (CS) community. New techniques, software, and computing resources are constantly being developed and published. Thus, there is a growing need to cite the latest foundational documents within our own papers.

For some time now the HEP community has had the tools to search for the latest papers for a given HEP topic [5–7]. Although this community has a demonstrated need to cite the latest-and-greatest ML documents, we currently do not have an efficient way of searching for them. Often an author would like to cite the latest research on a ML technique or the work where a common technique was originally proposed. These papers may not be commonly referenced or may be outside the researcher's area of expertise. This can lead to a long process of searching through similar domain specific papers, browsing the arXiv or even relying on a search engine. This is a tedious and error prone methodology.

Recently, the CERN Inter-Experimental LHC Machine Learning Working Group (IML) has sought to simplify this process by publishing a set of documents and/or websites which aggregate ML resources and papers. In particular, the living review [8] is a list of papers on a selection of ML topics. It is published both as a PDF document and as a website. This is a huge step forward from the previous methodology, but is still a manual, community supported list which captures a small snapshot of the state of machine learning at one time. Another tool made available by the IML group is the HEPML Resources website [9], which lists, for example workshops, tutorials, papers, and other information. Again, this is a helpful resource, but is supported by a small handful of individuals and is not regularly updated.

## 2 Ongoing Work and Opportunities for Improvement

For some time now the CMS collaboration has had a loose group of ML enthusiasts who met regularly, but without a formal structure. Recently, this changed with the formation of the CMS Machine Learning Group. One of the charges of this group is to provide an easy to access set of references, guides, tutorials, and a host of other analysis specific resources [10]. This work will both help to improve the IML resources, as CMS is one of the participating members of that group, as well as to provide CMS specific resources. Nevertheless, these are just the efforts of a single collaboration and cannot provide for the needs of an entire field of physics.

What is truly required in order to facilitate the needs of the HEP community is an automatically updated, cross referenced list of ML papers with topical tags. This could take the form of websites, much like iNSPIRE and the arXiv, or it could be something completely different. While we support the creation of such a resource, we admit that the details and implementation should be discussed further. We propose that these types of community discussions should take place in the course of the Snowmass 2021 process and that, at the very least, the initial groundwork for the resource be laid during this time.

## References


[1] Kim Albertsson et al. "Machine Learning in High Energy Physics Community White Paper". In: *Proceedings, 18th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2017)*. Vol. 1085. 2018, p. 022008. DOI: 10.1088/1742-6596/1085/2/022008. arXiv: 1807.02876 [physics.comp-ph].

[2] Alexander Radovic et al. "Machine learning at the energy and intensity frontiers of particle physics". In: *Nature* 560 (2018), p. 41. DOI: 10.1038/s41586-018-0361-2.

[3] Dan Guest, Kyle Cranmer, and Daniel Whiteson. "Deep Learning and its Application to LHC Physics". In: *Ann. Rev. Nucl. Part. Sci.* 68 (2018), p. 161. DOI: 10.1146/annurev-nucl-101917-021019. arXiv: 1806.11484 [hep-ex].

[4] Dimitri Bourilkov. "Machine and Deep Learning Applications in Particle Physics". In: *Int. J. Mod. Phys. A* 34 (2020), p. 1930019. DOI: 10.1142/S0217751X19300199. arXiv: 1912.08245 [physics.data-an].

[5] *arXiv*. 2020 (accessed August 28, 2020). URL: https://arxiv.org/.

[6] INSPIRE Team. *iNSPIRE*. (accessed August 28, 2020). URL: https://inspirehep.net/.

[7] *CERN Document Server*. 2020 (accessed August 28, 2020). URL: https://cds.cern.ch/.

[8] HEP ML Community. *A Living Review of Machine Learning for Particle Physics*. URL: https://iml-wg.github.io/HEPML-LivingReview/.

[9] Matthew Feickert et al. *iml-wg/HEP-ML-Resources: 2019*. Version 2019. Jan. 2020. DOI: 10.5281/zenodo.3626294. URL: https://doi.org/10.5281/zenodo.3626294.

[10] The CMS Collaboration. *CMS Machine Learning Documentation*. Unpublished. 2020 (accessed August 28, 2020). URL: https://cms-ml.github.io/documentation/.