

ML4HEP 2026

TIFR



TATA INSTITUTE
OF FUNDAMENTAL
RESEARCH

ML4HEP

4th Machine Learning for High Energy Physics School

STATISTICAL METHODS AND MACHINE LEARNING IN HIGH ENERGY PHYSICS

Today, High Energy Physics (HEP) research is at a crossroads. While the Large Hadron Collider (LHC) keeps accumulating data to establish the Standard Model on a solid footing, compelling theoretical underpinnings point to the existence of new physics at higher energy scales. In the future, the High Luminosity LHC will precisely measure the properties of the Higgs boson, using a few thousand petabytes of data. Many other precision experiments in HEP are under construction or going to start soon. Hence, the future course of the field will be largely data-driven.

Machine Learning techniques will be heavily employed in analyzing this humongous data for possible hints of new physics. Already, remarkable progress has been achieved in developing different classification, identification, characterization, and estimation strategies for use in the searches performed at LHC.

ORGANIZERS

Sunanda Banerjee	University of Wisconsin, USA
Satyaki Bhattacharya	SNP, India
Indumathi D	IMSc, India
Bhawna Gumber	University of Hyderabad, India
Partha Konar	PRL, India
Aruna Kumar Nayak	IOPB, India
Ritesh Kumar Singh	IISER, Kolkata, India

DATE

28 August 2023 - 8 September 2023

2024 in IOP
Bhubaneswar

2025 in IISER
Kolkata

Learning to discover: the Higgs boson machine learning challenge



Claire Adam-Bourdarios^a, Glen Cowan^b, Cécile Germain^c,
Isabelle Guyon^d, Balázs Kégl^{a,c}, David Rousseau^a

^a LAL, IN2P3/CNRS & University Paris-Sud, France

^b Physics Department, Royal Holloway, University of London, UK

^c TAO team, INRIA & LRI, CNRS & University Paris-Sud, France

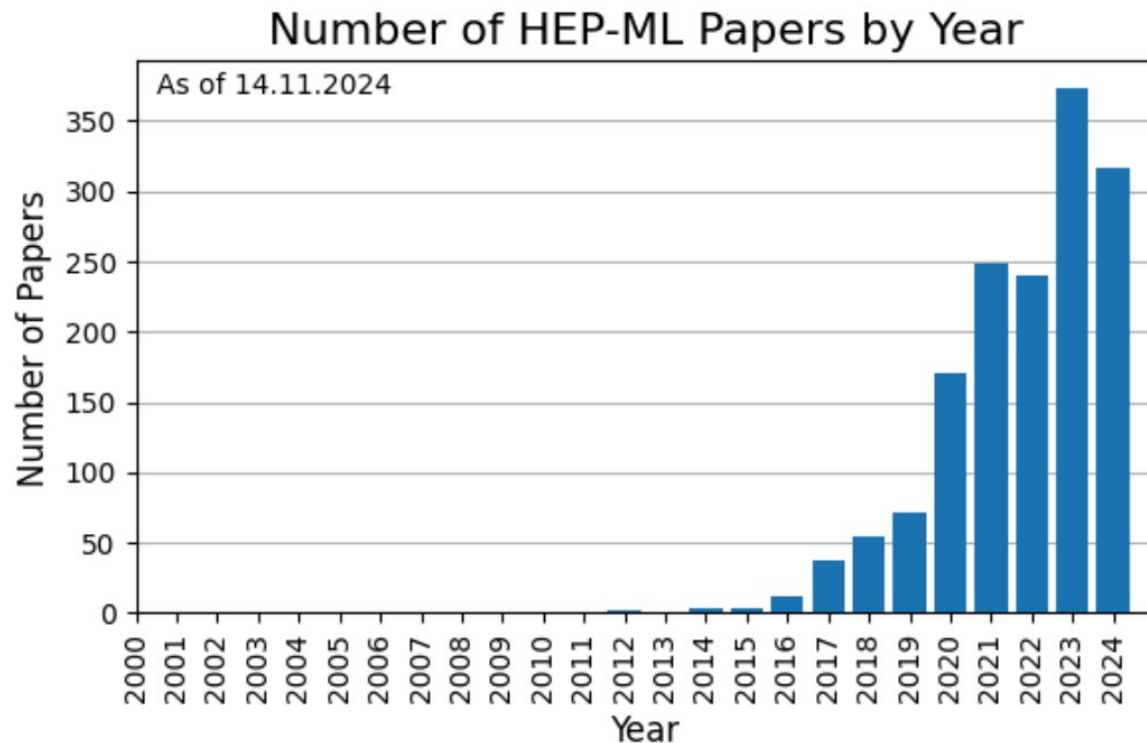
^d ChaLearn

21st January 2015, version 2.3

Preamble

The Higgs boson machine learning challenge (HiggsML or Challenge in short) has been set up to promote collaboration between high energy physicists and data scientists. The ATLAS experiment at CERN provided simulated data used by physicists to optimize the analysis of the Higgs boson. The Challenge is organized by a group of ATLAS physicists and data scientists. It was hosted by Kaggle at <https://www.kaggle.com/c/higgs-boson>; the challenge data is now available on <http://opendata.cern.ch/collection/ATLAS-Higgs-Challenge-2014>; This document provides technical background for the Challenge; reading and understanding it is not necessary to participate in the Challenge, but it is probably useful. No prior knowledge of high energy physics is required. This document has been minimally modified to serve as a permanent documentation for the dataset on opendata.cern.ch.

Explosion of ML in HEP



Machine Learning in Higgs Discovery

During the 1990s, ANNs became a standard data analysis tool within particle physics experiments of ever-increasing complexity. Highly sought-after fundamental particles, such as the Higgs boson, only exist for a fraction of a second after being created in high-energy collisions (e.g. $\sim 10^{-22}$ s for the Higgs boson). Their presence needs to be inferred from tracking information and energy deposits in large electronic detectors. Often the anticipated detector signature is very rare and could be mimicked by more common background processes. To identify particle decays and increase the efficiency of analyses, ANNs were trained to pick out specific patterns in the large volumes of detector data being generated at a high rate.

ANNs improved the sensitivity of searches for the Higgs boson at the CERN Large Electron-Positron (LEP) collider during the 1990s [44], and were used in the analysis of data that led to its discovery at the CERN Large Hadron Collider in 2012 [45]. ANNs were also used in studies of the top quark at Fermilab [46].