FERMILAB-PUB-18-436-ND
ACCEPTED

Title:
"Machine learning at the energy and intensity frontiers of particle physics"

Author List:

Alexander Radovic*
College of William & Mary
Mike Williams*
MIT
David Rousseau
LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France
Michael Kagan
SLAC National Accelerator Laboratory, Stanford CA, United States of America
Daniele Bonacorsi
University of Bologna
Alexander Himmel
Fermi National Accelerator Laboratory, Batavia IL, United States of America
Adam Aurisano
Cincinnati University
Kazuhiro Terao
SLAC National Accelerator Laboratory, Stanford CA, United States of America
Taritree Wongjirad
Tufts University

*Corresponding authors

Author Statement:

All authors contributed to writing this review. In addition, the corresponding authors served as the principal editors.

Preface

Our knowledge of the fundamental particles of nature and their interactions is elegantly summarized in the Standard Model (SM) of particle physics. Advancing our understanding in this field has required developing ever more advanced experiments which create extremely large and information rich data samples. The use of machine learning (ML) techniques is revolutionizing how we interpret these data samples, greatly increasing the discovery potential of present and future experiments. The use of ML at the frontiers of particle physics comes with unique challenges and opportunities, which this review will summarize.

Introduction

The SM is supported by an abundance of experimental evidence, and yet we know that it cannot be a complete theory of nature, for example since it cannot since it cannot incorporate gravity or explain dark matter. Furthermore, many properties of known particles, including neutrinos and the Higgs boson, have not yet been determined experimentally, and how the emergent properties of complex systems of fundamental particles arise from the underlying SM theory remains an enigma.

Many known particles were discovered using detectors that made subatomic particles visible to the human eye. For example, bubble chambers [1] were filled with superheated liquids that boiled when charged particles

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

passed through them. The particle paths were transformed into visible tracks of bubbles, which were photographed and then analyzed by large research teams. The LHC [2] detectors are much more complex and record data at far greater rates than was possible using bubble chambers. For example, the LHCb experiment [3] analyzes as many events every 6 seconds as the Big European Bubble Chamber recorded in its entire 11 years of running (1973–1983), and the data sets collected by the ATLAS and CMS experiments [4,5] are comparable to the largest industrial data samples. It is impossible for humans to visually inspect such vast amounts of data, and algorithms running on large computing farms took over long ago.

Over the past two decades, particle physics has been migrating towards the use of ML methods in the collection and analysis of its large data samples[6]. Pioneering studies employing neural networks (NNs) [7, 8] and boosted decision trees (BDTs) [8,9] at previous-generation experiments [11,12,13,14,15,16,17,18,19,20,21,22,23] laid the groundwork for the emergence of ML as an essential tool at the LHC. ML algorithms made important contributions to the discovery of the Higgs boson[24,25], and most data-analysis tasks now benefit from the use of ML. In parallel the field of ML has developed at a rapid pace, and in particular the new subfield of deep learning has delivered genuinely superhuman performance in a number of domains[26,27]. Incorporating these new tools while maintaining the scientific rigor required in particle physics analyses presents some unique challenges. This review focuses on the application and development of ML methods at the LHC, including recent advances based on deep learning. In addition, we present some examples of cutting-edge applications of deep learning within the sub-field of neutrino physics, where state-of-the-art methods, for example from computer vision, are naturally applicable.

Big Data at the LHC

The sensor arrays of the LHC experiments produce data at a rate of about a petabyte per second. Even after drastic data-reduction performed by the custom-built electronics used to readout the sensor arrays, which involves zero suppression of the sparse data streams and the use of various custom compression algorithms, the data rates are still too large to store indefinitely—as much as 50 terabytes per second, resulting in as much data every hour as Facebook collects globally in a year. This section first motivates why it is necessary to produce such immense data samples, before discussing how ML is being employed to more effectively select—in real time—which data to keep for further studies and which data to permanently discard. In addition, we show how the use of ML is leading to more efficient processing of these data using vast computing resources distributed around the world. Both of these Big Data challenges must be overcome before the LHC data can be used to advance our knowledge of fundamental particles.

LHC Operations

Einstein famously related mass to energy via $E = mc^2$ and, therefore, a powerful particle accelerator like the LHC, which is 27 km in circumference, is required to create particles orders of magnitude more massive than the proton, such as the Higgs boson. A Higgs boson is produced only once every few billion proton-proton collisions at the LHC. Many other interesting reactions occur orders of magnitude less often. To enable recording such data samples in a reasonable time frame, the LHC collides almost one billion protons per second.

High-energy collisions can produce hundreds of particles, and disentangling such complex events requires detectors with large and diverse sensor arrays. The ATLAS and CMS detectors each weigh approximately as much as the Eiffel tower, and contain roughly 100 million detection elements. Most of the particles produced in the LHC experiments decay before they can be detected by any of the sensors. Therefore, LHC analyses must infer what the underlying reactions were based on the properties of those particles which are detected. A wide variety of sensor technologies are used in the LHC detectors. The various signals from the particles detected by these sensor arrays are digitized, converting the physical processes involving subatomic particles into large collections of bytes. The extreme rate at which the LHC collides protons, along with the size and complexity of the LHC detectors, results in the production of such enormous data samples.

Real-time Analysis

The LHC experiments employ data-reduction schemes executed in real time, referred to as *triggers*, to identify what data to persist for future analysis and which data to permanently discard. For example, the ATLAS and CMS experiments each only keep about 1 in every 100 000 events, and yet their data samples are each still about 20 petabytes per year. The first step in deciding which events to keep relies on logic that is directly encoded into the hardware to enable the fastest possible decisions, such as devices known as Field Programmable Gate Arrays (FPGAs). ML is already used in this restrictive environment. For example, CMS employs ML in its trigger hardware to better estimate the momentum of muons [28], where the inputs to the algorithm are discretized to permit encoding the ML response in a large look-up table that is easily to programmed into the FPGAs.

In addition, the LHC experiments use huge computing farms to process their extreme data volumes searching for interesting signatures. In the case of LHCb, many of the reactions of greatest interest do not produce striking signatures in the detector, making it necessary to thoroughly analyze high-dimensional feature spaces in real time to efficiently classify events [29]. Since the first year of LHCb data collection, the primary algorithm used for such classification has been ML-based; specifically, a BDT was used for the first 2 years [30], which has since been replaced by a MatrixNet algorithm [31]. The use of ML is now ubiquitous, which has greatly improved the performance while satisfying the stringent robustness requirements of a system that makes irreversible decisions. Currently, 70% of all data persisted is classified by ML algorithms, and all charged-particle tracks are vetted by neural networks [32]. As an example of the impact of these ML methods, achieving the same sensitivity as a recent LHCb search for the dark matter analog of the photon, which was performed using data collected in 2016 [33], would have required collecting data for 10 years without the use of ML.

Actionable insights from computing metadata

Processing of the industrial-scale data samples collected by the LHC experiments is performed using the vast computing resources of the LHC Computing Grid, which are distributed across dozens of centers worldwide. The massive data volumes moved between Grid centres, and the large amount of CPU processing jobs used to access and analyse these data, generate an enormous amount of metadata information from which actionable insights can be extracted. ML techniques have recently begun to play a crucial role in increasing the efficiency of computing resource usage at the LHC [34,35,36] . One example is predicting which data will be accessed the most, as currently monitored by CMS [37] and LHCb [38], so that it becomes possible to optimise data storage at the Grid computing centres. Another example involves monitoring data transfer latencies over complex network topologies at CMS [39], using ML to identify problematic nodes and to predict likely congestions. Currently, ML informs the choices of the computing operations teams, but in the future it will be at the basis of fully automatic and adaptive models.

Machine Learning as an Established Tool in Particle Physics

After identifying and recording the most interesting LHC events and processing them on the Grid—two vital tasks supported by ML—the data are ready for exploration. The first step in interpreting these data involves grouping the signals recorded by various sensor elements according to which particle created them. The types and properties of the particles can then be inferred from the subsets of event information associated to them. Finally, after reconstructing all detected particles in the event, the data are analyzed to determine the underlying physical processes that created the particles. Interpreting such complex data samples is an extremely challenging task, which has been revolutionized by the use of ML techniques over the past years. About 2000 journal articles have been produced by the LHC experiments to date, providing a vast library of examples of the use of ML with these types of complex data sets. This section discusses a few highlights, including the role played by ML in the discovery of the Higgs boson [24, 25].

Determining particle properties

The use of ML to improve the determination of particle properties is now commonplace at all of the LHC experiments. For example, BDTs are used to increase the resolution of the CMS electromagnetic calorimeter [39]. When an electron or photon enters such a detector, it rapidly loses its energy which is subsequently collected and measured by the calorimeter. This deposited energy is often recorded by many different sensors, and the readings from these sensors must be clustered together to recover the original energy of the particle. Multivariate regression is used by CMS to train BDTs to provide corrections to these inferred energies based on all of the information contained in each calorimeter sensor. Applying these energy corrections to the decay of a Z boson into an electron-positron pair results in a substantial improvement of the mass resolution compared to the traditional clustering approach (see Fig. 1).

Discovery of the Higgs boson

As stated above, a Higgs boson is produced only once every few billion proton-proton collisions at the LHC; however, the Higgs boson usually decays in ways that mimic much more copiously produced processes. The cleanest experimental signature of the Higgs boson involves its decay into two muon-antimuon pairs, which occurs roughly once every 10 trillion proton-proton collisions. This and a few other processes were used in the Higgs discovery analyses. Most were selected due to their striking experimental signatures, which made it possible to obtain pure signals using relatively simple analyses. An important exception was the analysis of the Higgs boson decaying into two photons by the CMS experiment.

The CMS analysis involved searching for a small excess of diphoton candidates, manifest as a narrow peak in the diphoton mass spectrum, in the presence of a large smoothly distributed background. This background largely consisted of diphotons that originated from processes other than that of the Higgs decay, and from candidates formed from one real photon combined with an artificial photon signal, i.e. a photon inferred from the detector signals that did not correspond to an actual photon produced in the physical process. Two BDTs were used to improve the diphoton mass resolution by better determining which proton-proton collision the photons were produced in. Since both the SM signal Higgs process and the dominant background processes are well understood, it was possible to use simulated data samples to train a BDT, and based on the response of this BDT, the CMS diphotons were either discarded or kept for further analysis. The selected diphotons were also categorized using the BDT response, making it possible to analyze a rare—but highly pure—subset of Higgs decays separately. A simultaneous fit was performed to the mass distributions of all categories, which greatly enhanced the sensitivity to the presence of a Higgs signal. The increase in sensitivity due to the use of ML was equivalent to collecting 50% more data.

Determining the properties of the Higgs boson

The SM contains only one Higgs boson, which is the simplest explanation for the phenomenon known as electroweak symmetry breaking. Many extensions to the SM predict that there are many Higgs bosons, for example, supersymmetric theories predict a rich Higgs sector, while other theories predict that the Higgs itself is a composite object, i.e. that it is not a fundamental particle. The SM provides precise predictions for the properties of the Higgs, and it is vital that these predictions are tested experimentally to determine the nature of the Higgs particle discovered at the LHC.

The Higgs boson discovery analyses firmly established its interactions with the electroweak force-carrying particles, namely the photon, W and Z bosons. The SM also predicts that the Higgs interacts with fermions (quarks and leptons), and that the strength of each of these interactions is proportional to the fermion masses. This means that the Higgs is expected to decay into heavier quarks and leptons more often than into their lighter cousins. The ATLAS and CMS experiments have thus far observed the Higgs decaying into the heaviest kinematically accessible quark, the beauty quark [41, 42], and into the most massive lepton (a heavier version of the electron known as the τ). ML played a major role in each of these discoveries, though we will only describe the ATLAS search for H → $\tau^+\tau^-$ in detail here.

The study of τ particles is challenging because they decay before being detected and because their decays always involve neutrinos that escape detection carrying away energy. Furthermore, the decays of Z bosons to $\tau^+\tau^-$ occur about 1000 times more often often than the H → $\tau^+\tau^-$ signal. The ATLAS analysis divided the data sample into six distinct kinematic regions. A BDT was trained in each region using 12 weakly discriminating input features [43]. Figure 2 shows an example BDT response distribution obtained in one region. The combined analysis of all six regions provided strong evidence for the realization of the Higgs coupling to tau leptons in nature, with about 40% better sensitivity achieved through the use of ML. Thus far, the interactions of the Higgs with quarks and leptons appear to be consistent with the SM predictions. The accurate simulation leading to this result was eventually released through Kaggle as the basis of the 2014 Higgs Machine Learning Challenge [44], where data scientists competed to provide alternative ML methods to isolate the H → $\tau^+\tau^-$ signal.

A high-precision test of the Standard Model

The SM predicts that only three out of every billion B_s particles, which are bound states that contain a beauty quark, decay into a muon-antimuon final state. The fact that this decay rate is so highly suppressed in the SM contributes to it being extremely sensitive to potential quantum effects induced by as-yet-unknown particles, especially from an extended Higgs sector; for example certain supersymmetric theories predict an order of magnitude enhancement in this decay rate. The CMS and LHCb experiments were the first to find evidence for this decay using data samples collected in the first few years of the LHC, and a combined analysis of their data sets produced its first observation [45]. The analyses used BDTs to reduce the dimensionality of the feature space—excluding the mass—to one dimension, then an analysis was performed of the mass spectra across bins of BDT response. This approach preserved as much information as possible about the mass spectra of both the signal and backgrounds, providing the best possible sensitivity to this extremely rare decay of the B_s meson into a muon-antimuon final state. The decay rate observed is consistent with the SM prediction with a precision of about 25%, which places stringent constraints on many proposed extensions to the SM. Finally, a more recent update from LHCb achieved the first single-experiment observation [46], where achieving a similar sensitivity without the use of ML would have required collecting about four times more data. This is just one of many examples of high-precision tests of the SM at the LHC where ML has dramatically increased the power of the measurement.

The Emergence of Deep Learning as a Tool in Particle Physics

ML in particle physics, including the examples presented in the previous two sections, has traditionally involved the use of field-specific knowledge to engineer tools to extract the features of the data which were expected to be the most useful for a given measurement. This enables interpreting the incredibly rich initial data using only a small number of features. For example, in the B_s decay just discussed a human-designed tracking algorithm first reconstructs the paths taken by the muon and anti-muon in a magnetised particle physics detector, and from these paths the momenta of the particles are inferred. However, only the mass and the angle between the muon and antimuon are used in the BDT. The rest of the kinematic information is discarded.

For many tasks, information can be lost when these human designed tools are used to extract features that fail to fully capture the complexity of the problem. As in the fields of computer vision and natural language processing [27, 47], there is now a growing effort in particle physics to skip the feature-engineering step, and instead, use the full high-dimensional feature space to train cutting-edge ML algorithms, such as deep neural networks [48]. In this approach, domain expertise is used to design neural network architectures that are best suited to the specific problem. Studies of such applications have grown greatly within the last several years, beginning around 2014 with applications of deep neural networks to data analysis [49], then quickly expanding to the first applications of computer vision [50,51,52], and to the present broad study of deep learning throughout the field of particle physics [53,54,55,56].

This section highlights a few recent applications of two types of deep learning algorithms in particle physics: convolutional and recurrent neural networks (CNNs and RNNs, respectively) [57, 58]. The outputs of many particle physics detectors can be viewed as images, and the application of computer vision techniques is being explored in simplified settings by the LHC community [59,60,61,62,63,64,65] and with initial studies on ATLAS and CMS simulations [66,67]. However, the application of such techniques is more direct in the area of neutrino physics, for which reason we shall focus our discussion of CNNs to neutrino experiments. Similarly, there exist many applications of RNNs, but for the sake of brevity we will only discuss their use for the study of high-energy beauty quarks at ATLAS and CMS.

Computer vision for neutrino experiments

Loosely inspired by the structure of the visual cortex, CNNs employ a strategy that decreases their sensitivity to the absolute position of elements in an image and that makes them more robust against noise. Deep CNNs are able to extract complex features from images, and can now outperform humans in certain image-classification tasks. Another strength of CNNs is their ability to identify objects in an image, used for self-driving cars, thanks to translational invariant feature learning. This translational invariance presents a challenge for the LHC experiments, whose detectors consist of layers of distinct detector technologies moving out from the proton-proton collision region. These detectors provide rich information in the absolute reference frame of the detector, and innovative remapping is required to transform the data into a more natural format for a CNN-based approach. In contrast, this characteristic of CNNs is particularly useful for neutrino experiments, which necessarily employ large homogeneous detectors due to the incredibly small probability that a neutrino will interact within a small volume of material. Within these detectors, a neutrino interaction can take place anywhere, and locating them is a critical part of neutrino physics analyses.

The NOvA experiment [68] is filled with scintillating mineral oil, which emits light when a charged particle passes through it. Each NOvA event consists of two images: one taken from the top and the other from the side. NOvA has developed an ML algorithm [52] composed of two parallel networks inspired by the GoogLeNet [69] architecture, which extracts features from both views simultaneously and combines them to categorize neutrino interactions in the detector. This network, which improves the efficiency of selecting electron neutrinos by 40% with no loss in purity, has served as the event classifier in both an electron neutrino appearance search [70], and in a search for a new type of particle called a sterile neutrino [71].

The MicroBooNE experiment [71], which contains 90 tonnes of liquid argon, detects neutrinos sent from the booster neutrino beamline (BNB) at Fermilab. Each MircoBooNE event is a 33 megapixel image that likely contains background tracks caused by cosmic rays, and identifying neutrino interaction signals in such events, which vary in size from a few centimeters to meters, is one of the most challenging tasks in the experiment. MicroBooNE recently demonstrated the ability to detect neutrino interactions using a CNN [72]. Specifically, a Faster-RCNN [73] took advantage of spatially sensitive information from intermediate convolution layers to predict a bounding box that contains the secondary particles produced in a neutrino interaction. Figure 3 shows an example of output where the network successfully localized a neutrino interaction with high confidence. Finally, by taking advantage of accelerated computing on GPUs these CNNs can run significantly faster than the conventional algorithms used by previous neutrino experiments. This makes them ideally suited to the task of real-time image classification and object detection.

Recurrent neural networks for beauty quark identification

The study of high-energy beauty quarks is of great interest at the LHC, since these are frequently produced in the decays of Higgs bosons and top quarks, and predicted to be important components of the decays of supersymmetric and other hypothetical particles. A high-energy beauty quark radiates a substantial fraction of its energy in the form of a collimated stream of particles, called a jet, prior to forming a bound state with an antiquark or two additional quarks. This radiation is emitted over a distance comparable to the size of a proton, making it impossible to directly observe the emission process. The beauty-quark bound states only live for a picosecond, corresponding to mm to cm flight distances at the LHC, before randomly decaying into one of a thousand possible sets of commonly produced particles. Therefore, identifying jets that originate from high-energy beauty quarks relies on the ability to to determine whether particles were produced displaced from the proton-proton collision. Since jets typically contain between 10 and 50 particles, the number of potential discriminating features varies on a per-jet basis. Traditional jet identification algorithms rely on either explicitly identifying secondary production points from the crossing of particle trajectories, a highly challenging task, or compressing the features space using engineering and neglecting the correlations between particles when using single particle features. While combining such algorithms has been done with ML for some time [74,75], ML can also be used to improve identification using the low level particle features within a jet.

RNNs have proven to be extremely successful at processing long sequences of data, most famously acting as the new core of Google’s translation services [47]. RNNs process sequences in a way that information across the entire sequence can be accumulated and used. Applying an RNN to jet classification requires ordering the particles in the jet to form a sequence, for example ranking them by how incompatible they are with originating from the proton-proton collision. A set of features for each particle is provided to the RNN, which is trained to discriminate between beauty-quark jets and all other jet types. The use of an RNN at the ATLAS experiment reduced the misidentification rate by a factor of four relative to a non-ML-based algorithm [76]. When the RNN is itself used as an input feature in the subsequent training of a BDT or neural network, the misidentification rate is reduced by a factor of three relative to the ML result without the use of the RNN as an input feature [77]. Similar approaches are also actively explored by the CMS experiment [78], and more sophisticated RNN structures have been studied in a simplified setting which show promising results [79].

Training and Validation

The ML algorithms used in particle physics are typically trained using *supervised learning* [80] and data samples where the true origins, identities, and properties of the particles are known a priori. The algorithms learn to identify patterns in the training data, making it possible for them to predict information about particles in data samples where expert labeling of data is impossible. It is vital that any ML tool undergoes rigorous validation and testing, and that the uncertainty on its performance is well understood. There is always the possibility that some features used by an ML algorithm are not properly modeled in the training samples which—if not properly accounted for—could lead to a false discovery claim. Ultimately, we use ML tools to minimize uncertainties, and the validation procedures discussed in this section are important for gaining confidence in their behavior.

Learning from simulation

The need to understand what signals will look like in the detectors, and what other processes can mimic those signals, has led to the development of high-quality simulation tools. Furthermore, the SM provides accurate predictions of both the rates and kinematic distributions of many of the most challenging processes that can mimic interesting signatures, backgrounds, that contribute to particle physics data samples—providing important benchmarks for validating the simulation tools, and understanding their uncertainties. Therefore, simulated data samples are often used to train the ML algorithms since in such samples all information is known by construction. An important exception is that it is often possible to obtain highly pure background-only data samples, for example, by using events collected under different experimental conditions, and such samples are often used as background samples during training. A hybrid approach is also possible. The MicroBooNE CNN discussed above was trained using simulated neutrino interactions overlaid on top of cosmic-ray background images taken with the real detector.

Testing for bias

The quality and robustness of all ML tools are validated using well-known reactions recorded by the experiments. One approach, which is used by all LHC experiments [81,82], involves constructing special data samples in which the data are fully understood without the use of ML. For example, the LHCb experiment uses $J/\psi \rightarrow \mu^+\mu^-$ decays to validate its muon-identification neural network (μNN) [32]. The J/ψ is a copiously produced charm-anticharm bound state, which can be selected with 99.9% purity when applying the μNN to either the μ^+ or μ^-. Therefore, the identity of the other particle is known without using the μNN, providing an unbiased data sample where the μNN performance can be studied. Domain-specific knowledge is then employed to transfer what is learned on these validation samples, in terms of both the expected performance and its uncertainty, to any analysis that uses that specific ML algorithm. In the μNN case, the algorithm is studied in intervals of muon and event-level properties, and the detector response within these intervals is assumed to be independent of the process that produces the muon.

Another approach involves hybrid events, where the data are augmented with simulation to produce a test sample that mimics a signal. One example employed by NOvA [83] takes abundant and pure muon neutrino charged-current data and replaces the detected muon with a simulated electron. These hybrid events allow NOvA to study the performance of its ML algorithms on rare electron neutrino charged-current interactions, which are expected to look identical in the detector apart from the muon-to-electron swap. Similar techniques were used for H->tau+tau- by both ATLAS [43] and CMS[84].

The approaches presented in the previous paragraphs are reminiscent of the procedures used to characterise the performance of complex detectors in the past decades. Alternatively, tools developed by the ML community can be used to probe the response of the algorithms. For example, t-SNE [85] is a non-parametric embedding technique that allows one to visualize the proximity of points in a high-dimensional space using just two dimensions. It can be used to study the groupings of different events according to the features extracted by a deep neural network. Events which have overlapping extracted features, which the network interprets to be similar, are near each other in the t-SNE mapping. Conversely, events with little or no overlap are far from each other in the mapping. These t-SNE projections are used to ensure that the groupings match the intuition about the physical processes under study, to check if non-training events are embedded as expected, and even used in conjunction with auto-encoder neural networks to search for anomalies in large datasets. Figure 4 shows an example of such a t-SNE embedding using simulated neutrino interactions at the NOvA experiment.

Conclusions and outlook

Within the next decade the LHC will increase the rate at which it collides protons by an order of magnitude, resulting in much higher data rates and even more complex events to disentangle. Neutrino physics detectors will continue to increase in size and complexity. The tasks discussed in this review will become even more challenging in the future. Fortunately, the frontiers of ML are rapidly advancing, producing exciting new tools that are potentially applicable to a wide array of tasks in particle physics. By continuing to map the challenges faced in particle physics to those addressed by the ML community, it is then possible to turn the latest developments in ML into tools for discovery in high-energy particle physics, for example by conducting ML challenges with LHC benchmark data sets. A few potential future applications are briefly discussed below, which have already shown promising results on simplified test cases.

The ML community continues to discover powerful methods to process and classify complex data with inherent structure, such as trees and graphs. Complex data structures are prevalent at the LHC. The set of particles that make up a jet can be mapped to a tree structure. Above we demonstrated that RNNs can be used to identify jets that originate from beauty quarks, but this is just one of the many potential applications of RNNs, or graph convolutional networks, to the study of jets [86].

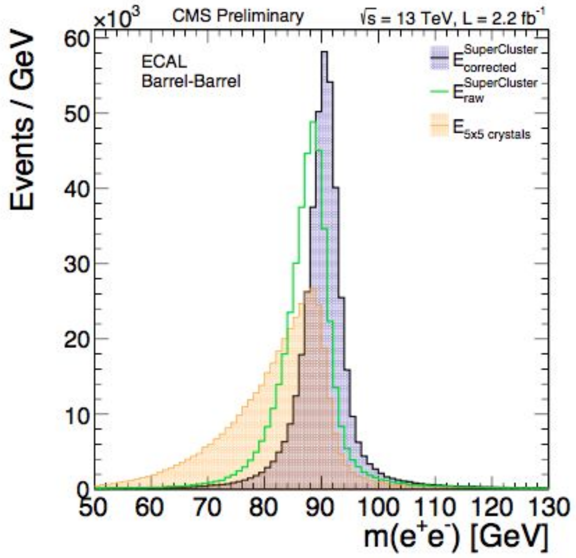
Generative models, which learn the probability distribution of features directly, are capable of producing high fidelity random data using cutting-edge tools such as Generative Adversarial Networks (GANs) [87] and Variational Auto-Encoders (VAE)[88,89]. A GAN employs one neural network to generate potential data samples using random noise as input, while a second network, the adversary, penalizes the generative network during training if the generated data can be distinguished from the training data. Though they are difficult to train, these networks can potentially generate large data samples much faster than existing simulation tools, offering the possibility of providing the orders of magnitude larger simulation samples that will be required by future experiments. Early work in this direction is encouraging [90,91,92], demonstrating that accurate simulations of a simplified calorimeter can be produced while achieving a dramatic decrease in required computational resources.

The adversarial approach can also be applied to training classifiers with the ability to enforce invariance to latent parameters. This represents a novel way to make classifiers robust against systematic uncertainties [93], and is a viable approach to avoid biasing a physical feature such as mass. A number of promising alternatives are also being investigated[94, 95, 96, 97], and some have even been deployed for analysis at LHCb[98]. All these approaches share the common theme of altering the training of the algorithms to reduce the potential bias learned. These are just a few of the exciting ML developments that are revolutionizing data interpretation in particle physics, greatly increasing the discovery potential of present and future experiments.

Table:

Figures:

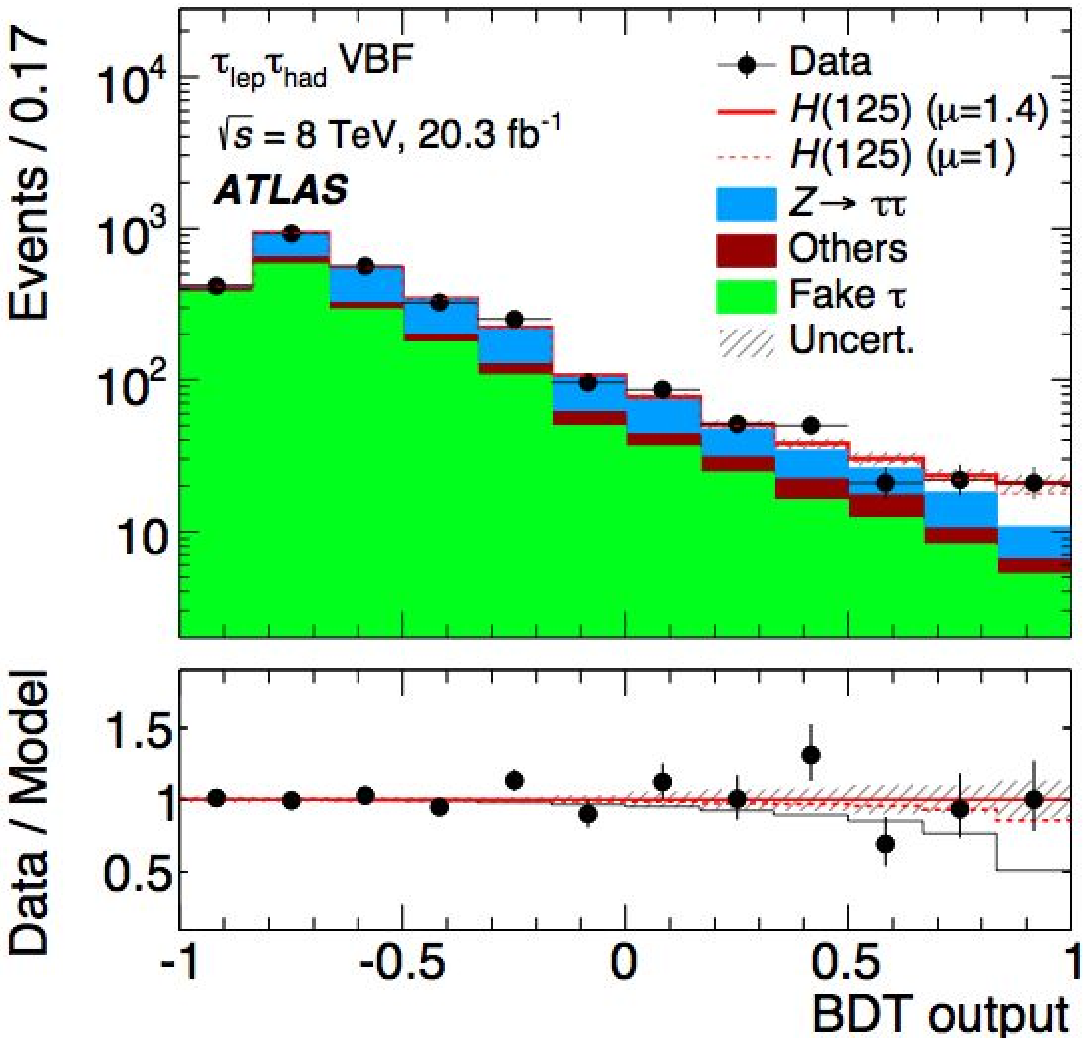
Figure 1, “Machine Learning for Calorimetry at CMS”:



Mass distribution of Z bosons decaying to electron pair measured in the central (barrel) part of the detector (for part of the 2015 data taking period, proton colliding energy $\sqrt{s}$=13 TeV and integrated luminosity of 2.2 fb^{-1}), whose true peak position is at 91 GeV, decaying into electron-positron pairs (yellow histogram) using only the raw electromagnetic calorimeter (ECAL) information, (green line) after clustering, and (blue histogram) after applying the ML-based corrections discussed in the text.

(*This figure is public as CMS-DP-2015/057 - https://twiki.cern.ch/twiki/bin/view/CMSPublic/EcalDPGResults*)

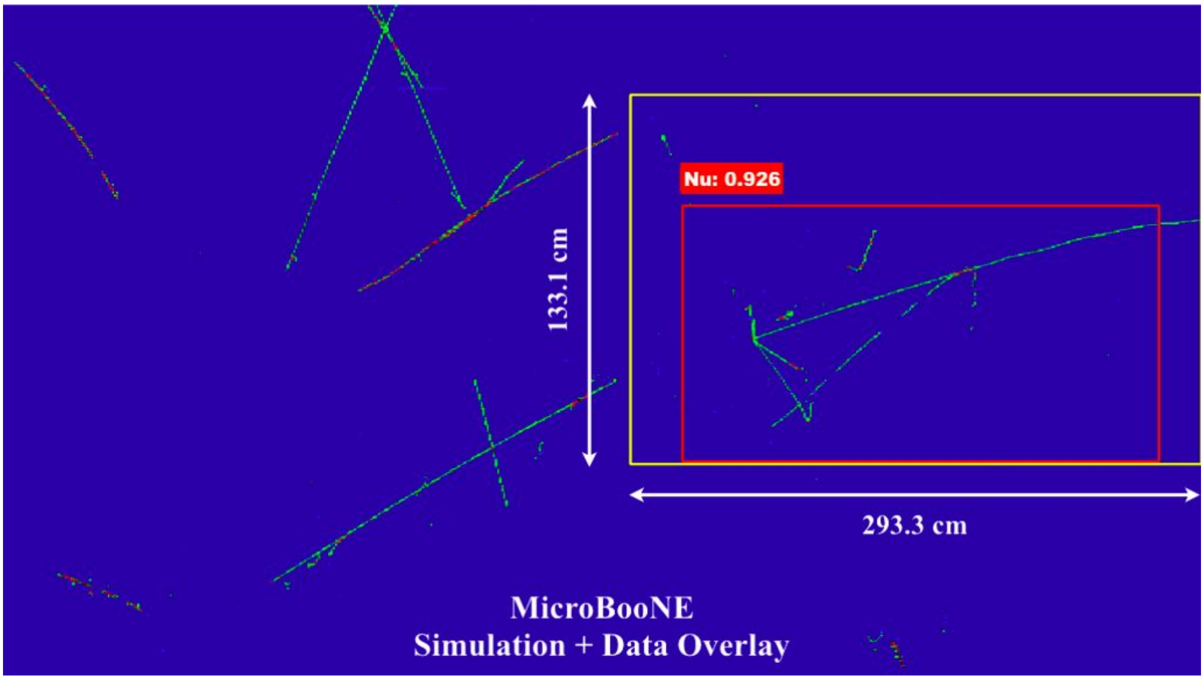
Figure 2, “Separating signal events from background in the ATLAS experiment”:



The figure shows the BDT score distribution for a search of the Higgs boson decaying in a τ lepton pair (H→ τ τ) in the ATLAS detector, for the 2012 data taking period (proton colliding energy $\sqrt{}$s=8 TeV and integrated luminosity of 20.3 fb^{-1}). The colored area shows the stacked contributions of the different background processes (Z→ τ τ another particle decaying also in a τ lepton pair , Fake τ where at least one τ lepton are misidentified, and other smaller contributions) while the black dots show the data. The dotted red line shows the expected excess in mainly the rightmost two intervals attributed to the Higgs boson for production rate identical to the Standard Model expectation (μ=1), while the solid red line shows the fitted excess (μ=1.4), still compatible with the Standard Model . The bottom plot show the ratio of data to the fitted model.

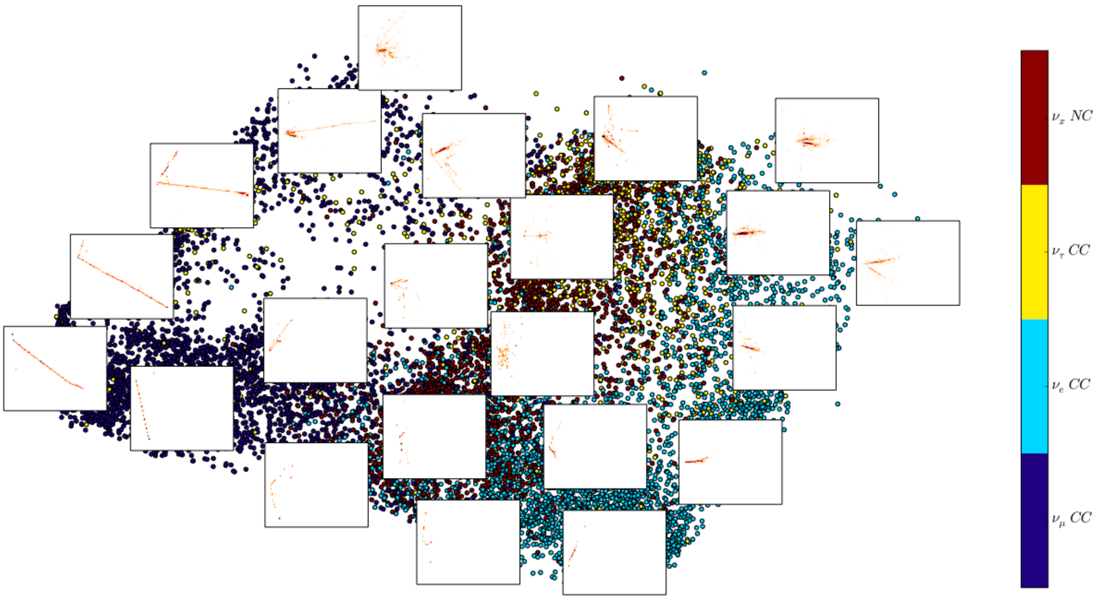
(*This figure was published in [26]*)

Figure 3, “Neutrino Selection and Isolation in MicroBooNE ”:



MicroBooNE event display showing a simulated neutrino interaction overlaid onto a cosmic-ray background image taken using the real detector. The yellow box, which is drawn as a reference and based on simulation information, contains all charge depositions caused by secondary charged particles produced in the simulated neutrino interaction. The CNN receives as input the image without the yellow box, and draws the red box that successfully captures the most interesting part of the neutrino interaction.

Figure 4, “Exploring the NOvA Event Selection Neural Network Using t-SNE”:



Projection of the extracted features of a NOvA neutrino interaction CNN into a two-dimensional space using the t-SNE method. The points represent events from the CNN training sample, where the colors denote the true event type (see vertical legend at right). The subplots show example event topologies from those points in the t-SNE space. One can see that the various event types are clustered into distinct regions in the horizontal direction, while event activity is found to be correlated with the event locations in the vertical direction.

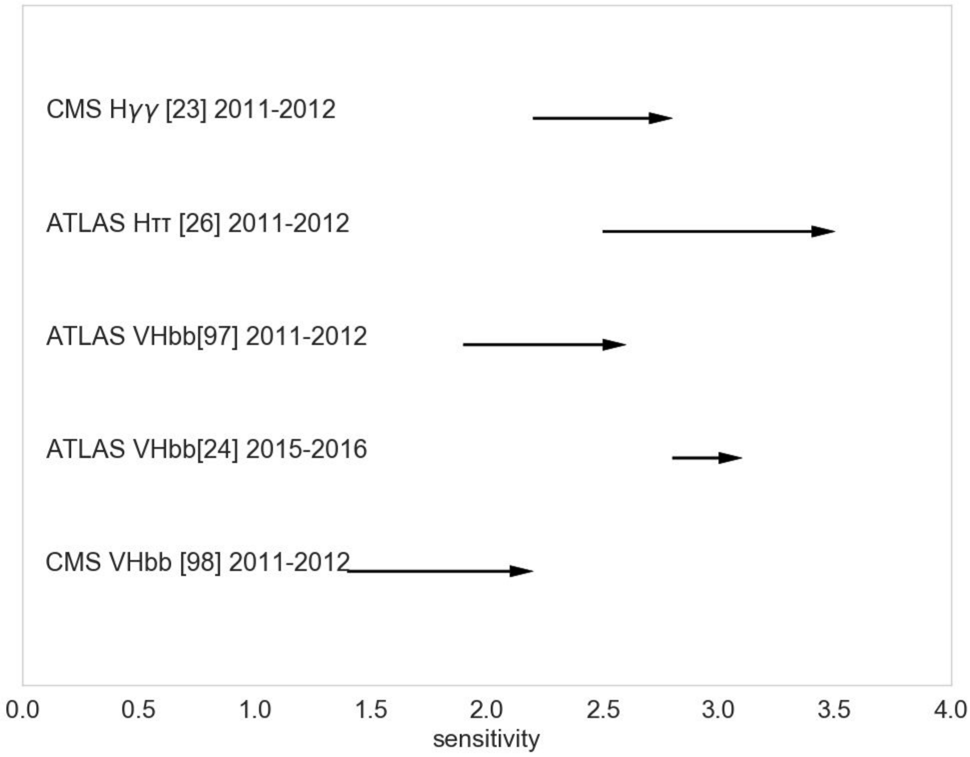
This figure is blessed for public use by the NOvA collaboration, but has not previously been in a publication.

Boxes:

Impact of Machine Learning on the discovery and study of the Higgs Boson
TABLE AND TEXT REPLACED BY FOLLOWING PLOT AND TEXT

Analysis	Data collection year	No ML sensitivity | p-value	ML sensitivity | p-value	Relative data gain | factor
CMS $H \rightarrow \gamma\gamma$ [25]	2011-2012	2.2 | 0.014	2.7 | 0.0035	51% | 4.0
ATLAS $H \rightarrow \tau^+\tau^-$[43]	2011-2012	2.5 | 0.0062	3.4 | 0.00034	85% | 18
ATLAS VH→bb[99]	2011-2012	1.9 | 0.029	2.5 | 0.0062	73% | 4.7
ATLAS VH→bb [41]	2015-2016	2.8 | 0.0026	3.0 | 0.00135	15% | 1.9
CMS VH→bb[100]	2011-2012	1.4 | 0.081	2.1 | 0.018	125% | 4.5

Five key decay modes of the Higgs boson where ML greatly increased the sensitivity of the LHC experiments. For each analysis, the sensitivity both without and with ML is given, and both the p-values and the equivalent number of Gaussian standard deviations are provided (here we only present analyses that provided both ML and non-ML results, most recent analyses only report the superior ML-based results). The increase in sensitivity achieved using ML, as measured by the ratio of p-values, ranges from about 2 to 20. An alternative figure of merit is the additional amount of data that would need to be collected to reach the ML-based sensitivity without using ML, which varies from 15% to 125%.



Five key decay modes of the Higgs boson where ML greatly increased the sensitivity of the LHC experiments. For each analysis, the arrow shows the gain in sensitivity (in number of Gaussian standard deviation) from using ML (here we only present analyses that provided both ML and non-ML results, most recent analyses only report the superior ML-based results). An alternative figure of merit is the additional amount of data that would need to be collected to reach the ML-based sensitivity without using ML, which varies from 15% to 125%.

References:

[1] Glaser, D. A. Some Effects of Ionizing Radiation on the Formation of Bubbles in Liquids. Phys.Rev. 87 665-665. (1952)
[2] Evans, L., & Bryant, P. LHC Machine, JINST 3 S08001. (2008)
[3] Alves, A.Augusto, Jr. et al. The LHCb Detector at the LHC. JINST 3 S08005 (2008)
[4] Aad, G. et al.The ATLAS Experiment at the CERN Large Hadron Collider. JINST 3 S08003 (2008)
[5] Chatrchyan, S. et al. The CMS Experiment at the CERN LHC. JINST 3 S08004 (2008)
[6] Bhat, P. Multivariate Analysis Methods in Particle Physics. Annual Review of Nuclear and Particle. Science Vol. 61:281-309 (2011)
[7] Rosenblatt, F..Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books (1961).
[8] Reed, R. & Marks, R. Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks. A Bradford book. MIT Press. (1999)
[9] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. Classification and regression trees. Wadsworth international group. (1984).
[10] Freund, Y. and Schapire,R. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55 nº 1, 119 (1997).

[11]Barate, R. et al. Determination of |V(ub)| from the measurement of the inclusive charmless semileptonic branching ratio of b hadrons. Eur. Phys. J. C6, 555 (1999).
[12] Akers, R. et al. A measurement of the production of D*±mesons on the Z^0 resonance Z. Phys C 67 27 (1995).
[13]Chiappetta, P. , Colangelo, P. , De Felice, P., Nardulli, G. & Pasquariello, G. Higgs search by neural networks at LHC. Phys. Lett. B322, 219 (1994).
[14]Peterson, C. , Rognvaldsson, T. & Lonnblad, L. JETNET 3.0: A Versatile artificial neural network package. Comput. Phys. Commun. 81, 185 (1994).
[16] Buskulic, D. et al. Measurement of the tau polarisation at the Z resonance. Phys C 59, 369 (1993).
[17] Babbage, W. S. & Thompson, L. F. The Use of neural networks in gamma - pi0 discrimination. Nucl. Instrum. Meth. A330, 482 (1993).
[18] Lonnblad, L., Peterson, C. & Rognvaldsson, T. Pattern Recognition in High Energy Physics with Artificial Neural Networks - JETNET 2.0. Computer Physics Communications 70, 167 (1992).
[19]Peterson, C. , Rögnvaldsson, T. An introduction to artificial neural networks. In 1991 CERN School of Computing, ed. C. Verkerk, CERN Report 92/02 (1992).
[20] Lonnblad, L., Peterson, C., & Rognvaldsson, T. Using neural networks to identify jets. Nucl. Phys. B349, 675 (1991).
[21] Lonnblad, L., Peterson, C. & Rognvaldsson, T. Finding Gluon Jets With a Neural Trigger. Phys. Rev. Lett. 65, 1321 (1990).
[22] Denby, B. H. Neural Networks and Cellular Automata in Experimental High-energy Physics. Comput. Phys. Commun. 49, 429 (1988).
[23] Roe, B. et al. Boosted Decision Trees as an Alternative to Artificial Neural Networks for Particle Identification, Nucl.Instrum.Meth. A543 (2005)
[24] Aad, G. et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Phys. Lett. B716, 1(2012).
[25] Chatrchyan, S. et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. B716, 30–61 (2012).
[26] Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. Nature 529 484–489 (2016)
[27] Russakovsky, O. et al. ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision (IJCV) 115 211-252. (2015)
[28] Carnes, A.M., et al. BDTs in the Level 1 Muon Endcap Trigger at CMS. Proceedings of 18th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (2017)
[29] Aaij, R. et al. The LHCb Trigger and its Performance in 2011. JINST 8, P04022 (2013).
[30] Gligorov, V. V. & Williams, M. Efficient, Reliable and fast high-level triggering using a bonsai boosted decision tree. JINST 8, P02013 (2013). **A Bonsai BDT which identifies data in real time at LHCb, and is used by over 200 journal articles to date.**
[31] Likhomanenko, T., Ilten, P., Khairullin, E., Rogozhnikov, A., Ustyuzhanin, A. & Williams, M. LHCb Topological Trigger Reoptimization, Proceedings, 21st International Conference on Computing in High Energy and Nuclear Physics (CHEP 2015): Okinawa, Japan, April 13-17, 2015, J. Phys. Conf. Ser. 664, 082025 (2015).
[32] Aaij, R. et al. LHCb Detector Performance. Int. J. Mod. Phys. A30, 1530022 (2015).
[33] Aaij, R. et al. Search for dark photons in 13 TeV pp collisions, Phys. Rev. Lett. 120, 061801 (2018).
[34] Hushchyn, M. et al. GRID Storage Optimization in Transparent and User-Friendly Way for LHCb Datasets J. Phys: Conf. Series 898, 062023 (2017).
[35] Derkach, D. et al., LHCb trigger streams optimization, J. Phys: Conf. Series 898, 062026 (2017).
[36] Borisyak, M., Ratnikov, F., Derkach, D. & Ustyuzhanin, A. Towards automation of data quality system for CERN CMS experiment, J. Phys. Conf. Ser. 898 092041 (2017).
[37] Kuznetsov, V. et al. Predicting dataset popularity for the CMS experiment. J. Phys. Conf. Ser. 762, 012048 (2016).
[38] Hushchyn, M., Charpentier, P. & Ustyuzhanin, A. Disk Storage Management for LHCb Based on Data Popularity Estimator, J. Phys. Conf. Ser. 664, 042026 (2015).

[39] Bonacorsi, D., Diotalevi, T., Magini, N., Sartirana, A., Taze, M. & Wildish, T. Monitoring data transfer latency in CMS computing operations, Proceedings of CHEP'15, J. Phys. Conf. Ser. 664, 032033 (2015).
[40] CMS Collaboration. Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s}$ = 7 TeV, JINST 8, P09009 (2013).
[41] Aaboud, M. et al. Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector, JHEP 1712, 024 (2017.)
[42] CMS Collaboration, Evidence for the decay of the Higgs Boson to Bottom Quarks, CMS-PAS-HIG-16-044 (2017).
[43] Aad, G. et al. Evidence for the Higgs- boson Yukawa coupling to tau leptons with the ATLAS detector. JHEP 04, 117 (2015).
[44] Adam-Bourdarios, C. et al. The Higgs boson machine learning challenge. JMLR **42** 19 (2014). **Helped popularise particle physics in the general ML community, and advertised recent advances in ML within the particle physics community.**
[45] Khachatryan, V. et al. Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data. Nature 522, 68 (2015).
[46] Aaij, R. et al. Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays. Phys. Rev. Lett. 118, 191801 (2017).
[47] Yonghui, W. et al. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. arXiv:1609.08144 (2016)
[48] LeCun, Y., Bengio, Y., & Hinton, G. Deep learning. Nature,521,436-444 (2015)
[49] Baldi, P., Sadowski, P. & Whiteson, D. Searching for Exotic Particles in High-Energy Physics with Deep Learning,, Nature Commun. 5, 4308 (2014). **This paper started the use of "DL" in HEP, and walks through the difference between shallow networks with high level features, and deep networks that find their own high level features.**
[50] de Oliveira, L., Kagan, M., Mackey, L., Nachman, B. & Schwartzman, A. Jet-Images — Deep Learning Edition. JHEP 1607, 069 (2016)**. Started the investigation of DL approaches to QCD Jet work, with detailed discussion of CNNs and supporting exploration of the network behavior.**
[51] Racah, E. et al. Revealing Fundamental Physics from the Daya Bay Neutrino Experiment using Deep Neural Networks, IEEE ICMLA 2016, 892 (2016)
[52] Aurisano, A. et al. A Convolutional Neural Network Neutrino Event Classifier. J. Instrum. 11, P09001 (2016). **First convolutional neural network used for a physics analysis[70], with a detailed discussion of the method, and comparison to more traditional identification methods.**
[53] Sadowski, P, et al. Efficient Antihydrogen Detection in Antimatter Physics by Deep Learning. Journal of Physics Communications, Volume 1, Number 2, (2017)
[54] Renner, J., Background rejection in NEXT using deep neural networks. JINST 12 no.01, T01004 (2017)
[55] Wielgosza, M., Skoczen, A., &´Mertik, M. Using LSTM recurrent neural networks for monitoring the LHC superconducting magnets. Nucl.Instrum.Meth. A867 40-50 (2017)
[56] Edelen, A.L. et al. Neural Networks for Modeling and Control of Particle Accelerators. IEEE Trans.Nucl.Sci. 63 no.2, 878-897. (2016)
[57] LeCun, Y. et al. Backpropagation applied to handwritten zip code recognition. Neural Comput.,1,541-551 (1989)
[58] F.A. Gers, J. Schmidhuber, and F. Cummins. Neural Computation - 2000. Learning to forget: Continual prediction with lstms 12:2451-2471. (1999)
[59] Cogan, J., Kagan, M., Strauss, E. & Schwartzman, A. Jet-Images: Computer Vision Inspired Techniques for Jet Tagging, JHEP 1502, 118 (2015).
[60] Baldi, P., Bauer, K., Eng, C., Sadowski, P. & Whiteson, D. Jet substructure classification in high-energy physics with deep neural networks. Phys.Rev. D93, 094034 (2016).
[61] Barnard, J., Dawe, E.N., Dolan, M.J. & Rajcic, N. Parton shower uncertainties in jet substructure analyses with deep neural networks. Phys.Rev. D95, 014018 (2017).
[62] Komiske, P.T., Metodiev, E.M. & Schwartz, M.D. Deep learning in color: towards automated quark/gluon jet discrimination. JHEP 1701, 110 (2017) .

[63] de Oliveira, L., Paganini, M. & Nachman, B. Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis. Comput. Softw. Big. Sci. 1 (2017).
[64] Kasieczka, G., Plehn, T., Russell, M. & Schell, T. Deep-learning top taggers or the end of QCD? JHEP 05, 006 (2017)
[65] Shimmin, C., Sadowski, P., Baldi, P., Weik, E., Whiteson, D., Goul, E. & Søgaard, A. Decorrelated Jet Substructure Tagging using Adversarial Neural Networks. Phys. Rev. D 96, 074034 (2017)
[66] The ATLAS Collaboration. Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector. ATL-PHYS-PUB-2017-017.
[67] R.Barate et al. New Developments for Jet Substructure Reconstruction in CMS. CMS-DP-2017-027.
[68] Ayres, D.S. et al. The NOvA Technical Design Report. FERMILAB-DESIGN-2007-01 (2007)
[69] Szegedy, C. et al. Going Deeper with Convolutions. IEEE CVPR 1 (2015).
[70] Adamson, P. et al. Constraints on Oscillation Parameters from ν_e Appearance and ν_μ Disappearance in NOvA. Phys. Rev. Lett. 118, 231801 (2017)
[71] Adamson, P. et al. Search for active-sterile neutrino mixing using neutral-current interactions in NOvA. Phys. Rev. D96, 072006 (2017).
[71] Acciarri, R. et al. Design and Construction of the MicroBooNE Detector. JINST 12 no.02, P02017. (2017)
[72] Acciarri, R. et al. Convolutional neural networks applied to neutrino events in a liquid argon time projection chamber. J. Instrum. 12, P03011 (2017).
[73] Ren, S., He, K., Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE TPAMI 29, 061137 (2017).
[74] ATLAS Collaboration. Performance of b-jet identification in the ATLAS experiment. J. Instr. 11, P04008 (2016).
[75] CMS Collaboration. CMS Phase 1 heavy flavour identification performance and Developments. CERN-CMS-DP-2017-013.
[76] ATLAS Collaboration. Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment. ATL-PHYS-PUB-2017-003.
[77] ATLAS Collaboration. Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run. ATL-PHYS-PUB-2017-013.
[78] CMS Collaboration. Heavy flavor identification at CMS with deep neural networks. CMS-DP-2017-005
[79] Guest, D., Collado, J., Baldi, P. Hsu, S., Urban, G. & Whiteson, D. Jet Flavor Classification in High-Energy Physics with Deep Neural Networks. Phys.Rev. D94, 112002 (2016).
[80] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. Learning representations by back-propagating errors, Nature 323 533–536+. (1986)
[81] Aaboud, M. et al. Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data. Eur. Phys J. C77, 195 (2017).
[82] CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at sqrt(s) = 8 TeV. JINST 10, P06005 (2015).
[83] Sachdev, K. Muon Neutrino to Electron Neutrino Oscillation in NOvA. FERMILAB-THESIS-2015-20 (2015)
[84] Chatrchyan S. et al., Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons, JHEP 1405, 104 (2014).
[85] van der Maaten, L.J.P. Accelerating t-sne using tree- based algorithms. Journal of Machine Learning Research 15, 3221(2014).
[86] Louppe, G., Cho, K., Becot, C. & Cranmer, K. QCD- Aware Recursive Neural Networks for Jet Physics. (2017), arXiv:1702.00748.
[87] Goodfellow, I. J. et al. Generative Adversarial Networks. ANIPS 27, 2672 (2014).
[88] Rezende, D.J., Mohamed, S. & Wierstra, D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models, ICML 32,II1278 (2014).
[89] Kingma, D.P. & Welling, M. Auto-Encoding Variational Bayes. arXiv:1312.6114, Proceedings, International Conference on Learning Representations (ICLR), Banff (2014)

[90]de Oliveira L., Paganini M. and Nachman B. Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis, Computing and Software for the Big Science. 1 no.1, 4 (2017)
[91] Paganini, M. de Oliveira, L. & Nachman, B. CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks. Phys. Rev. Lett. 120, 042003 (2018)
[92] Carminati, F., et al. Calorimetry with Deep Learning: Particle Classification, Energy Regression, and Simulation for High-Energy Physics. Deep Learning for Physical Sciences Workshop at NIPS. (2017)
[93] Louppe, G., Kagan, M. & Cranmer, K. Learning to Pivot with Adversarial Networks. ANIPS 30, 6699 (2017).
Part of a wave of work that essentially constitutes a more nuanced loss function. Along with similar work with BDTs, detailed in the frontier section of this review, it could be a new paradigm for training ML tools in HEP.
[94] Stevens, J. & Williams, M. uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers. JINST 8, P12013 (2013).
[95] Rogozhnikov, A., Bukva, A., Gligorov, V. V., Ustyuzhanin, A. & Williams, M. New approaches for boosting to uniformity. JINST 10, T03002 (2015).
[96] Dery, L.M., Nachman, B., Rubbo, F. & Schwartzman, A. Weakly supervised classification in high energy physics. JHEP 2017, 45 (2017).
[97] Baldi, P., Cranmer, K., Faucett, T., Sadowski, P. & Whiteson, D. Parameterized neural networks for high-energy physics. Eur. Phys. J. C 76, 235 (2016).
[98] Aaij, R. et al. Search for hidden-sector bosons in $B^0 \to K^{*0}\mu^+\mu^-$ decays. Phys. Rev. Lett. 115, 161802 (2015).
[99] Aad, G. et al. Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated (W/Z)H production with the ATLAS detector. JHEP 01, 069 (2015)
[100] Chatrchyan, S. et al. Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks. Phys. Rev. D89, 012003 (2014)