REVIEW

Machine learning at the energy and intensity frontiers of particle physics

Alexander Radovic[1]*, Mike Williams[2]*, David Rousseau[3], Michael Kagan[4], Daniele Bonacorsi[5,6], Alexander Himmel[7], Adam Aurisano[8], Kazuhiro Terao[4] & Taritree Wongjirad[9]

Our knowledge of the fundamental particles of nature and their interactions is summarized by the standard model of particle physics. Advancing our understanding in this field has required experiments that operate at ever higher energies and intensities, which produce extremely large and information-rich data samples. The use of machine-learning techniques is revolutionizing how we interpret these data samples, greatly increasing the discovery potential of present and future experiments. Here we summarize the challenges and opportunities that come with the use of machine learning at the frontiers of particle physics.

The standard model of particle physics is supported by an abundance of experimental evidence, yet we know that it cannot be a complete theory of nature because, for example, 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 the way in which the emergent properties of complex systems of fundamental particles arise from the underlying standard-model theory remains unknown.

Many known particles were discovered using detectors that made subatomic particles visible to the human eye. For example, bubble chambers[1] filled with superheated liquids that boil when charged particles pass through them transform the paths of the particles into visible tracks of bubbles, which can then be photographed and analysed. The detectors at the Large Hadron Collider (LHC)[2] are much more complex and record data at far greater rates than is possible using bubble chambers. For example, the LHCb experiment[3] analyses as many events every six seconds as the Big European Bubble Chamber recorded in its entire 11 years of operation (1973–1983), and the datasets collected by the ATLAS[4] and CMS[5] experiments at the LHC are comparable to the largest industrial data samples. It is impossible for humans to visually inspect such large amounts of data; algorithms running on large computing farms took over this task long ago.

Over the past two decades, particle physics has been migrating towards the use of machine-learning methods in the collection and analysis of its large data samples[6]. Pioneering studies that used neural networks[7,8] and boosted decision trees (BDTs)[9,10] in previous-generation experiments[11–22] laid the groundwork for the emergence of machine learning as an essential tool at the LHC. Machine-learning algorithms made important contributions to the discovery of the Higgs boson[23,24] and most data-analysis tasks now benefit from the use of machine learning. In parallel, the field of machine learning has developed at a rapid pace and, in particular, the subfield of deep learning has delivered superhuman performance in several domains[25,26]. Incorporating these tools while maintaining scientific rigour required in particle-physics analyses presents new challenges. This Review focuses on the application and development of machine-learning methods at the LHC, including recent advances based on deep learning. In addition, we present some example applications of deep learning within the subfield of neutrino physics, in which state-of-the-art methods, such as 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 one petabyte per second. Even after drastic data reduction 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 the data indefinitely—as much as 50 terabytes per second, resulting in as much data every hour as Facebook collects globally in a year[27]. In this section we first motivate why it is necessary to produce such immense data samples, before discussing how machine learning is being used 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 machine learning 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 m to energy E via $E = mc^2$, where c is the speed of light in a vacuum. A powerful particle accelerator such as the LHC, which is 27 km in circumference, is therefore 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 such data samples to be recorded in a reasonable time frame, the LHC collides nearly 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 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 on the basis of the properties of the particles that are detected. A wide variety of sensor technologies are used in the LHC detectors. The various signals from the particles that are detected by these sensor arrays are digitized, converting the physical processes involving subatomic particles into large collections of bytes.

[1]College of William and Mary, Williamsburg, VA, USA. [2]Massachusetts Institute of Technology, Cambridge, MA, USA. [3]LAL, Université Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France. [4]SLAC National Accelerator Laboratory, Menlo Park, CA, USA. [5]Università di Bologna, Bologna, Italy. [6]INFN Sezione di Bologna, Bologna, Italy. [7]Fermi National Accelerator Laboratory, Batavia, IL, USA. [8]University of Cincinnati, Cincinnati, OH, USA. [9]Tufts University, Medford, MA, USA. *e-mail: aradovic@wm.edu; mwill@mit.edu

The extreme rate at which the LHC collides protons, along with the size and complexity of the LHC detectors, results in the production of enormous data samples.

Real-time analysis

The LHC experiments use data-reduction schemes executed in real time, referred to as triggers, to identify which data to retain for future analysis and which to permanently discard. For example, the ATLAS and CMS experiments each keep only about 1 in every 100,000 events. Despite this, 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 encoded directly into the hardware to enable the fastest possible decisions, such as into devices known as field-programmable gate arrays (FPGAs). Machine learning is already used in this environment; for example, CMS uses machine learning in its trigger hardware to better estimate the momentum of muons[28], with the inputs to the algorithm discretized to enable the machine-learning response to be encoded in a large look-up table that is easily programmed into the FPGAs.

In addition, the LHC experiments use huge computing farms to process the extreme volumes of data and search for interesting signatures. In the case of the LHCb experiment, many of the reactions of greatest interest do not produce striking signatures in the detector, making it necessary to thoroughly analyse 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 machine-learning-based; specifically, a BDT was used for the first two years[30], which has since been replaced by a MatrixNet algorithm[31]. The use of machine learning is now ubiquitous, which has greatly improved performance while satisfying the stringent robustness requirements of a system that makes irreversible decisions. Currently, 70% of all data retained are classified by machine-learning algorithms and all charged-particle tracks are vetted by neural networks[32]. As an example of the effect of these machine-learning methods, achieving the same sensitivity as a recent LHCb search for the dark-matter analogue of the photon, which was performed using data collected in 2016[33], would have required 10 years of data collection without the use of machine learning.

Actionable insights from computing metadata

Processing of the industrial-scale data samples collected by the LHC experiments is performed using the computing resources of the LHC Computing Grid, which are distributed across dozens of centres worldwide. The massive volumes of data moved between grid centres, and the large number 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. Machine-learning techniques have recently begun to play a crucial part in increasing the efficiency of computing-resource usage at the LHC[34–36]. 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 optimize data storage at the grid centres. Another example involves monitoring data-transfer latencies over complex network topologies at CMS[39], using machine learning to identify problematic nodes and to predict likely congestions. Currently, machine learning informs the choices of the computing-operations teams, but in the future it form the basis of fully automatic and adaptive models.

Machine learning as an established tool

After identifying and recording the most interesting LHC events and processing them on the Computing Grid—two vital tasks supported by machine learning—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 with them. Finally, after reconstructing all detected particles in the event, the data are analysed to determine the underlying physical processes that created the particles.
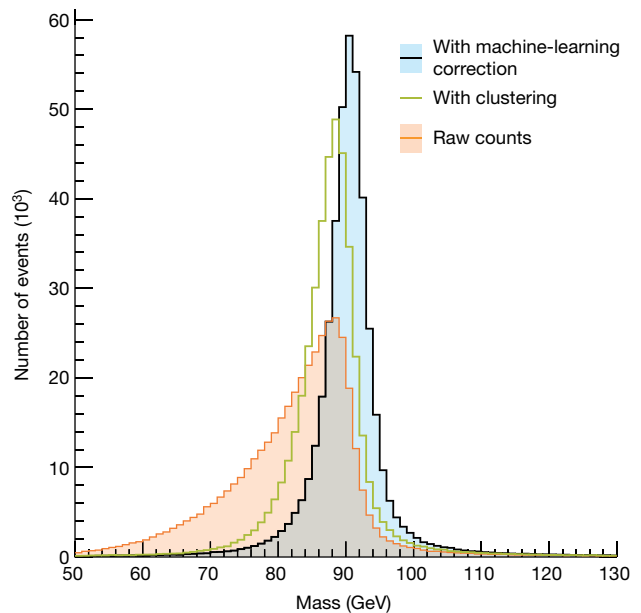

Fig. 1 | Machine learning for calorimetry at CMS. The mass distribution of Z bosons that decay to electron–positron pairs ($Z \rightarrow e^+e^-$), as measured in the central part of the CMS detector and binned into 1-GeV bins, is shown for three cases: using only the raw information from the detector (orange), after clustering the data (green) and after applying the machine-learning-based corrections discussed in the text (blue). The true position of the peak for this decay is 91 GeV. Image adapted from ref. [101] under a CC BY 4.0 license, copyright CERN, reused with permission.

Interpreting such complex data samples is an extremely challenging task, which has been revolutionized by the use of machine-learning techniques. About 2,000 journal articles have been produced by the LHC experiments to date, providing a large library of examples of the use of machine learning with these types of complex dataset. In this section we discuss a few highlights, including the role of machine learning in the discovery of the Higgs boson[23,24].

Determining particle properties

The use of machine learning 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[40]. 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 on the basis of 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 in 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 owing 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.
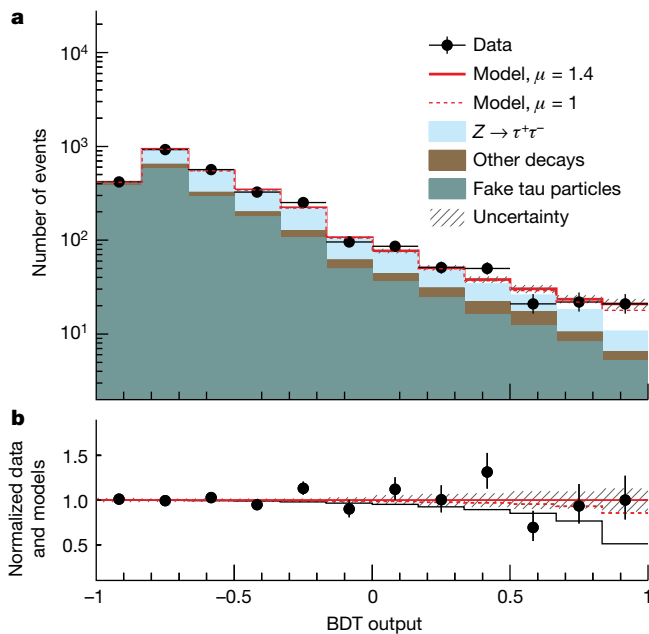
Fig. 2 | Separating signal events from background in the ATLAS experiment. a, The BDT-score distribution for a search for the Higgs boson decaying to a tau-lepton pair ($H \rightarrow \tau^{+}\tau^{-}$), with bin widths of 0.17. The black circles show the score of a machine-learning algorithm known as a BDT for data from the ATLAS detector during the 2012 data-taking period, where the error bars show the Poisson statistical uncertainties. This BDT was trained to distinguish a Higgs signal from various non-Higgs backgrounds. The coloured area shows the stacked contributions of the different background processes: $Z \rightarrow \tau^{+}\tau^{-}$ decays (blue), other particles decaying to a tau-lepton pair (brown) and fake tau particles (where at least one tau lepton is misidentified; green). The dotted red line shows the expected total counts assuming a Higgs-boson production rate identical to the standard-model expectation ($\mu = 1$); the solid red line shows the expected total counts assuming a Higgs-boson production rate of $\mu = 1.4$, which is still compatible with the standard model; the hatched area shows the systematic uncertainty on the expected total count. The excess counts compared to the coloured region (mainly in the rightmost two bins) are attributed to the Higgs boson. **b**, The ratios of the data (black circles), expected counts for $\mu = 1$ (dashed red line), expected counts for $\mu = 1.4$ (solid red line) and uncertainty (hatched region) from **a** to the expected counts for $\mu = 1.4$ (so the solid red line is identically 1) are shown, along with the ratio of the expected counts excluding the Higgs contribution (the sum of the green, brown, blue and hatched regions in **a**) to the expected counts for $\mu = 1.4$ (black line). Image adapted from ref. [43] under a CC BY 4.0 license, copyright CERN.

The CMS analysis involved searching for a small excess of diphoton candidates, manifested 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 the Higgs decay and from candidates formed from one real photon combined with an artificial photon signal (that is, 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. Because both the standard-model Higgs process and the dominant background processes are well understood, it was possible to use simulated data samples to train a BDT. On the basis of the response of this BDT, the CMS diphotons were either discarded or kept for further analysis. The diphotons selected were also categorized using the BDT response, making it possible to analyse 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 machine learning was equivalent to collecting 50% more data.

Determining the properties of the Higgs boson

The standard model contains only one Higgs boson, which is the simplest explanation for the phenomenon known as electroweak symmetry breaking. Many extensions to the standard model predict that there are many Higgs bosons; for example, super-symmetric theories predict a rich Higgs sector, and other theories predict that the Higgs boson is a composite object, not a fundamental particle. The standard model provides precise predictions for the properties of the Higgs boson 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 boson and Z boson. The standard model also predicts that the Higgs boson interacts with fermions (quarks and leptons) and that the strength of each of these interactions is proportional to the masses of the fermions. This means that the Higgs boson 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 boson 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 tau lepton). Machine learning had a major role in each of these discoveries, although we describe only the ATLAS search for the decay of the Higgs boson into an antitau–tau pair ($H \rightarrow \tau^{+}\tau^{-}$) in detail here.

The study of tau particles is challenging because they decay before being detected and because their decays always involve neutrinos that escape detection and carry away energy. Furthermore, the decay $Z \rightarrow \tau^{+}\tau^{-}$ occurs about 1,000 times more often than does $H \rightarrow \tau^{+}\tau^{-}$. 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]. In Fig. 2 we show 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 boson coupling to tau leptons in nature, with about 40% better sensitivity achieved through the use of machine learning. Thus far, the interactions of the Higgs boson with quarks and leptons appear to be consistent with standard-model predictions. The simulation that led 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 machine-learning methods to isolate the $H \rightarrow \tau^{+}\tau^{-}$ signal. Table 1 shows the impact of machine learning on the measurement of several key processing involving Higgs bosons at ATLAS and CMS.

A high-precision test of the standard model

The standard model predicts that only three out of every billion B_s particles—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 standard model 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 super-symmetric 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 these datasets produced the first observation of it[45]. The analyses used BDTs to reduce the dimensionality of the feature space—excluding the mass—to one dimension and 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 standard-model prediction with a precision of about 25%, which places stringent constraints on many proposed extensions to the standard model. Finally, a more recent update from the LHCb experiment achieved the first single-experiment observation[46]; achieving a similar sensitivity without the use of machine

Table 1 | Effect of machine learning on the discovery and study of the Higgs boson

Analysis	Years of data collection	Sensitivity without machine learning	Sensitivity with machine learning	Ratio of P values	Additional data required
CMS[24] $H \rightarrow \gamma\gamma$	2011–2012	2.2σ, $P = 0.014$	2.7σ, $P = 0.0035$	4.0	51%
ATLAS[43] $H \rightarrow \tau^+\tau^-$	2011–2012	2.5σ, $P = 0.0062$	3.4σ, $P = 0.00034$	18	85%
ATLAS[99] $VH \rightarrow bb$	2011–2012	1.9σ, $P = 0.029$	2.5σ, $P = 0.0062$	4.7	73%
ATLAS[41] $VH \rightarrow bb$	2015–2016	2.8σ, $P = 0.0026$	3.0σ, $P = 0.00135$	1.9	15%
CMS[100] $VH \rightarrow bb$	2011–2012	1.4σ, $P = 0.081$	2.1σ, $P = 0.018$	4.5	125%

Five key measurements of three decay modes of the Higgs boson H for which machine learning greatly increased the sensitivity of the LHC experiments, where V denotes a W or Z boson, γ denotes a photon and b a beauty quark. For each analysis, the sensitivity without and with machine learning is given, in terms of both the P values and the equivalent number of Gaussian standard deviations σ. (We present only analyses that provided both machine-learning-based and non-machine-learning-based results; the more recent analyses report only the machine-learning-based results.) The increase in sensitivity achieved by using machine learning, as measured by the ratio of P values, ranges roughly from 2 to 20. An alternative figure of merit is the minimal amount of additional data that would need to be collected to reach the machine-learning-based sensitivity without using machine learning, which varies from 15% to 125%.

learning would have required the collection of about four times as much data. This is just one of many examples of high-precision tests of the standard model at the LHC for which machine learning has markedly increased the power of the measurement.

The emergence of deep learning

Machine learning 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 that are expected to be the most useful for a given measurement. This enables the incredibly rich initial data to be interpreted using only a small number of features. For example, in the aforementioned B_s decay, a human-designed tracking algorithm first reconstructs the paths taken by the muon and the antimuon in a magnetized particle-physics detector, and from these paths the momenta of the particles are inferred. However, only the dimuon mass and the angle between them 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[26,47], there is 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 machine-learning 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 substantially in number and complexity within the past several years, beginning around 2014 with applications of deep neural networks to data analysis[49], quickly expanding to the first applications of computer vision[50–52] and to the current broad study of deep learning throughout the field of particle physics[53–56].

In this section we highlight a few recent applications of two types of deep learning algorithm 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[59–65] and with initial studies on ATLAS and CMS simulations[66,67]. However, such techniques are more naturally applicable in the area of neutrino physics, for which reason we focus our discussion of CNNs to neutrino experiments. Similarly, there are many applications of RNNs, but for brevity we discuss only 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 use a strategy that decreases their sensitivity to the absolute position of elements in an image and that makes them more robust to noise. Deep CNNs are able to extract complex features from images and now outperform humans in certain image-classification tasks. Another strength of CNNs is their ability to identify objects in an image, as demonstrated for example by their use in self-driving cars, owing to translation-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, which is transformed into a more natural format for a CNN-based approach. By contrast, this characteristic of CNNs is particularly useful for neutrino experiments, which necessarily use large homogeneous detectors owing to the incredibly small probability that a neutrino will interact within a small volume of material. A neutrino interaction can take place anywhere within these detectors and locating them is a critical part of neutrino-physics analyses.

The detectors of the NOvA experiment[68] are 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. The NOvA collaboration has developed a machine-learning algorithm[52] composed of two parallel networks inspired by the GoogleNet[69] architecture. The NOvA CNN 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 searches both for the appearance of electron neutrinos[70] and for a new type of particle called a sterile neutrino[71].

The detector at the MicroBooNE experiment[72], which contains 90 tonnes of liquid argon, detects neutrinos sent from the booster neutrino beamline at Fermilab. Each MircoBooNE event corresponds to a 33-megapixel image that probably contains background tracks caused by cosmic rays. Identifying signals of neutrino interactions in such events, in which both the signal and background tracks vary in size from a few centimetres to metres, is one of the most challenging tasks of the experiment. MicroBooNE recently demonstrated the ability to detect neutrino interactions using a CNN[73]. Specifically, an algorithm called Faster-RCNN[74] uses spatially sensitive information from intermediate convolution layers to predict a bounding box that contains the secondary particles produced in a neutrino interaction. In Fig. 3 we show an example output in which the network successfully localized a neutrino interaction with high confidence. Finally, by taking advantage of accelerated computing on GPUs, these CNNs can run much 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.

RNNs for beauty–quark identification

The study of high-energy beauty quarks is of great interest at the LHC because these particles are frequently produced in the decays of Higgs bosons and top quarks and are predicted to be important components of the decays of super-symmetric 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, before 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 observe the emission process directly. The beauty-quark bound states live for only a picosecond, corresponding to millimetre-to centimetre-scale flight distances at the LHC, before randomly decaying into one of a thousand possible sets of commonly produced particles. Therefore, to identify jets that originate from high-energy beauty quarks, it is necessary to be able to determine whether particles were produced directly in the proton–proton collision or in the subsequent decay of a long-lived bound state at a location displaced
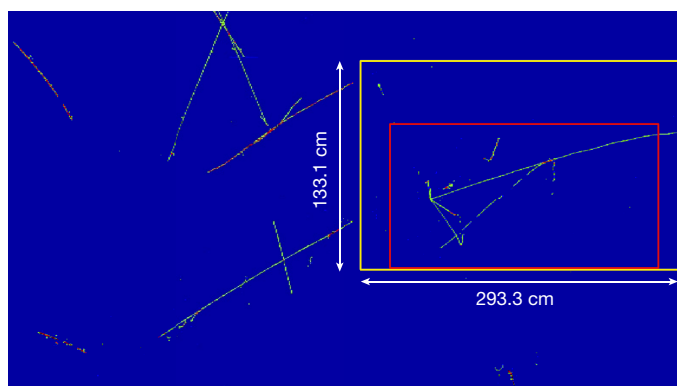
Fig. 3 | Neutrino selection and isolation in MicroBooNE. The MicroBooNE event display shows a simulated neutrino interaction (inside the yellow box) overlaid on a cosmic-ray background image taken using the real detector. The yellow boxed region contains all charge depositions caused by secondary charged particles being produced in the simulated neutrino interaction. The CNN receives as input the display without the yellow box indicated and draws the red box, which matches the yellow box remarkably well and successfully captures the most interesting part of the neutrino interaction. Image adapted from ref. [73], copyright Sissa Medialab, reused with permission of IOP Publishing.

from the proton–proton collision. Because 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 identifying secondary production points explicitly from the crossing of particle trajectories, a highly challenging task, or compressing the information with engineered features and neglecting the correlations between particles when using single-particle features. Although such algorithms have been combined with machine learning for some time[75,76], machine learning could also be used to improve identification by 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 core of Google's current translation service[47]. RNNs process sequences in such a way that information across the entire sequence can be accumulated and used. Applying an RNN to jet classification requires the particles in the jet to be ordered to form a sequence, such as by 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 types of jet. The use of an RNN at the ATLAS experiment reduced the misidentification rate by a factor of four relative to a non-machine-learning-based algorithm[77]. When the RNN is itself used as an input feature in the subsequent training of a BDT or neural network, the misidentification rate was reduced by a factor of three relative to the machine-learning result without the use of the RNN as an input feature[78]. Similar approaches are also being explored at the CMS experiment[79]; more sophisticated RNN structures have been studied in a simplified setting and show promising results[80].

Training and validation

The machine-learning algorithms used in particle physics are typically trained using supervised learning[81] and data samples for which 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 for which expert labelling of data is impossible. It is vital that any machine-learning 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 a machine-learning algorithm are not properly modelled in the training samples, which—if not properly accounted for—could lead to a false discovery. Ultimately, we use machine-learning tools to minimize uncertainties; the validation

procedures discussed in this section are important for gaining confidence in the behaviour of these tools.

Learning from simulation

The need to understand what signals will look like in the detectors and what other processes can mimic the signals has led to the development of high-quality simulation tools. Furthermore, the standard model provides accurate predictions of the rates and kinematic distributions of many of the processes that can mimic interesting signatures (referred to as backgrounds) and 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 machine-learning algorithms because 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, such as 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 cosmic-ray background images taken with the real detector.

Testing for bias

The quality and robustness of all machine-learning tools are validated using well-known reactions recorded by the experiments. One approach, which is used by all LHC experiments[82,83], involves constructing data samples in which the data are fully understood without the use of machine learning. For example, the LHCb experiment uses $J/\psi \rightarrow \mu^+\mu^-$ decays to validate its muon-identification neural network (μNN)[32]; J/ψ is a copiously produced charm–anticharm bound state, which can be identified with 99.9% purity when basing a selection criterion on the μNN response to either the antimuon (μ^+) or muon (μ^-). The identity of the other particle is therefore known without using the μNN, providing an unbiased data sample with which the performance of the μNN can be studied. Domain-specific knowledge is then used 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 machine-learning algorithm. In the case of the μNN, the algorithm is studied in ranges of the values of muon and event-level properties and the response of the detector within these ranges is assumed to be independent of the process that produces the muon.

Another approach involves hybrid events, whereby the data are augmented with simulations to produce a test sample that mimics a signal. One example used by NOvA[84] takes abundant and pure muon-neutrino charged-current data and replaces the detected muon with a simulated electron. These hybrid events allow the performance of NOvA's machine-learning algorithms to be studied on rare electron-neutrino charged-current interactions, which are expected to look identical to muon-neutrino charged-current interactions in the detector apart from the muon-to-electron swap. Similar techniques were used for the $H \rightarrow \tau^+\tau^-$ decay by ATLAS[43] and CMS[85].

The approaches presented above are reminiscent of the procedures used to characterize the performance of complex detectors in past decades. Alternatively, tools developed by the machine-learning community can be used to probe the response of the algorithms. For example, t-distributed stochastic neighbour embedding (t-SNE)[86] is a non-parametric embedding technique that allows the proximity of points in a high-dimensional space to be visualized using only two dimensions. It can be used to study the groupings of different events according to the features extracted by a deep neural network. Events with 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 being studied, to check whether non-training events are embedded as expected, and can even be used in conjunction with auto-encoder neural networks to search for anomalies
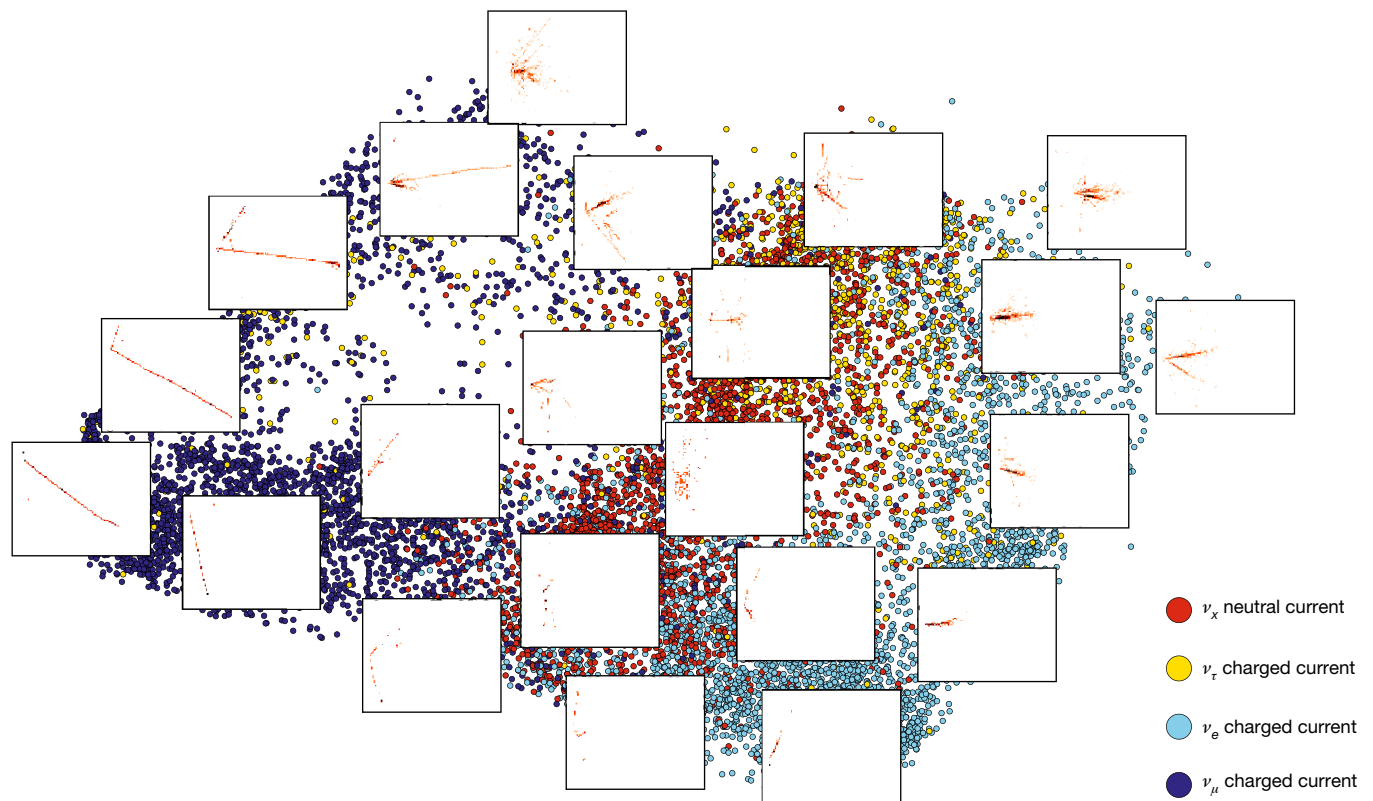
Fig. 4 | Exploring NOvA's event-selection neural network using t-SNE. The features extracted using NOvA's neutrino-interaction CNN are projected into a two-dimensional space using the t-SNE method. The points represent events from the CNN training sample, with the colours denoting the true event types: muon-neutrino (ν_μ) charged-current interactions (dark blue), electron-neutrino (ν_e) charged-current interactions (light blue), tau-neutrino (ν_τ) charged-current interactions (yellow) and various neutrino (ν_x) neutral-current interactions (red). The subplots show example event topologies from points in the two-dimensional t-SNE space, with the intensity of the colour indicating the amount of energy deposited and the axes denoting the spatial location of the charge deposits in the detector. The various types of event are clustered into distinct regions in the horizontal direction, while the multiplicity of the particles in each event is found to be correlated with the location of the events in the vertical direction.

in large datasets. In Fig. 4 we provide an example of such a t-SNE 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. Fortunately, machine learning is advancing rapidly, producing 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 machine-learning community, it is possible to turn the latest developments in machine learning into tools for discovery in high-energy particle physics, such as by conducting machine-learning competitions with LHC benchmark datasets (https://www.kaggle.com/c/trackml-particle-identification). We briefly discuss a few potential future applications below, which have already shown promising results for simplified test cases.

The machine-learning community continues to discover powerful methods for processing and classifying 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. We have already discussed how 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 of graph convolutional networks, to the study of jets[87].

Generative models, which learn the probability distribution of features directly, are capable of producing simulated data that closely approximate experimental data using tools such as generative adversarial networks[88] and variational auto-encoders[89,90]. A generative adversarial network uses one neural network, the 'generator', to generate potential data samples using random noise as input, while a second network, the 'adversary', penalizes the generator during training if the data that it generates can be distinguished from the training data. Although they are difficult to train, these networks can potentially generate large data samples much faster than can 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[63,91,92], demonstrating that accurate simulations of a simplified calorimeter can be produced while achieving a marked decrease in the computational resources required.

The adversarial approach can also be applied to training classifiers with the ability to enforce invariance to latent parameters. This represents a new way of making classifiers robust against systematic uncertainties[93] and is a viable approach to avoid biasing a physical feature such as mass[65]. Several promising alternatives are also being investigated[94–97], some of which have been used for analysis at LHCb[98]. All of these approaches share the common theme of altering the training of the algorithms to reduce the potential bias learned. These are only a few of the machine-learning developments that are revolutionizing data interpretation in particle physics, greatly increasing the discovery potential of present and future experiments.

1. Glaser, D. A. Some effects of ionizing radiation on the formation of bubbles in liquids. *Phys. Rev.* **87**, 665 (1952).
2. Evans, L. & Bryant, P. LHC machine. *J. Instrum.* **3**, S08001 (2008).
3. Alves, A. A. Jr et al. The LHCb detector at the LHC. *J. Instrum.* **3**, S08005 (2008).

4. Aad, G. et al. The ATLAS experiment at the CERN Large Hadron Collider. *J. Instrum.* **3**, S08003 (2008).
5. Chatrchyan, S. et al. The CMS experiment at the CERN LHC. *J. Instrum.* **3**, S08004 (2008).
6. Bhat, P. Multivariate analysis methods in particle physics. *Annu. Rev. Nucl. Part. Sci.* **61**, 281–309 (2011).
7. Rosenblatt, F. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms* (Spartan Books, Berlin, 1961).
8. Reed, R. & Marks, R. *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks* (MIT Press, Cambridge, 1999).
9. Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. *Classification and Regression Trees* (Wadsworth International Group, Belmont, 1984).
10. Freund, Y. & Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **55**, 119–139 (1997).
11. The ALEPH Collaboration. Determination of $|V_{ub}|$ from the measurement of the inclusive charmless semileptonic branching ratio of b hadrons. *Eur. Phys. J. C* **6**, 555–574 (1999).
12. OPAL Collaboration. A measurement of the production of $D^{*\pm}$ mesons on the Z^0 resonance. *Z. Phys. C* **67**, 27–44 (1995).
13. Chiappetta, P., Colangelo, P., De Felice, P., Nardulli, G. & Pasquariello, G. Higgs search by neural networks at LHC. *Phys. Lett. B* **322**, 219–223 (1994).
14. Peterson, C., Rognvaldsson, T. & Lönnblad, L. JETNET 3.0—a versatile artificial neural network package. *Comput. Phys. Commun.* **81**, 185–220 (1994).
15. Buskulic, D. et al. Measurement of the tau polarisation at the Z resonance. *Z. Phys. C* **59**, 369–386 (1993).
16. Babbage, W. S. & Thompson, L. F. The use of neural networks in γ-π^0 discrimination. *Nucl. Instrum. Methods A* **330**, 482–486 (1993).
17. Lönnblad, L., Peterson, C. & Rognvaldsson, T. Pattern recognition in high energy physics with artificial neural networks — JETNET 2.0. *Comput. Phys. Commun.* **70**, 167–182 (1992).
18. Peterson, C. & Rögnvaldsson, T. S. An introduction to artificial neural networks. In *14th CERN School of Computing* (ed. Verkerk, C.) 113–170 (CERN, 1992).
19. Lönnblad, L., Peterson, C. & Rögnvaldsson, T. Using neural networks to identify jets. *Nucl. Phys. B* **349**, 675–702 (1991).
20. Lönnblad, L., Peterson, C. & Rögnvaldsson, T. Finding gluon jets with a neural trigger. *Phys. Rev. Lett.* **65**, 1321–1324 (1990).
21. Denby, B. Neural networks and cellular automata in experimental high-energy physics. *Comput. Phys. Commun.* **49**, 429–448 (1988).
22. Roe, B. P. et al. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nucl. Instrum. Methods A* **543**, 577–584 (2005).
23. 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. B* **716**, 1–29 (2012).
24. Chatrchyan, S. et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B* **716**, 30–61 (2012).
25. Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
26. Russakovsky, O. et al. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
27. Vagata, P. & Wilfong, K. Scaling the Facebook data warehouse to 300 PB. *Facebook Code* https://code.fb.com/core-data/scaling-the-facebook-data-warehouse-to-300-pb/ (2014).
28. CMS Collaboration. Boosted decision trees in the level-1 muon endcap trigger at CMS. In *18th International Workshop on Advanced Computing and Analysis Techniques in Physics Research* 21–25 (CERN, 2017).
29. Aaij, R. et al. The LHCb trigger and its performance in 2011. *J. Instrum.* **8**, P04022 (2013).
30. Gligorov, V. V. & Williams, M. Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree. *J. Instrum.* **8**, P02013 (2013).
    This paper presents a boosted decision tree that identifies data in real time at LHCb and that has been used in more than 200 journal articles so far.
31. Likhomanenko, T. et al. LHCb topological trigger reoptimization. *J. Phys. Conf. Ser.* **664**, 082025 (2015).
32. The LHCb Collaboration. LHCb detector performance. *Int. J. Mod. Phys. A* **30**, 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. Ser.* **898**, 062023 (2017).
35. Derkach, D. et al. LHCb trigger streams optimization. *J. Phys. Conf. Ser.* **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. et al. Monitoring data transfer latency in CMS computing operations. *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. *J. Instrum.* **8**, P09009 (2013).
41. The ATLAS Collaboration. Evidence for the $H \to bb^-$ decay with the ATLAS detector. *J. High Energy Phys.* **12**, 24 (2017).
42. CMS Collaboration. *Evidence for the Decay of the Higgs Boson to Bottom Quarks*. Report No. CMS-PAS-HIG-16-044, https://cds.cern.ch/record/2278170 (CERN, 2017).
43. Aad, G. et al. Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector. *J. High Energy Phys.* **4**, 117 (2015).
44. Adam-Bourdarios, C. et al. The Higgs boson machine learning challenge. *J. Mach. Learn. Res. Worksh. Conf. Proc.* **42**, 19–55 (2014).
    This work helped to popularize particle physics in the general machine-learning community and advertised recent advances in machine learning within the particle-physics community.
45. CMS Collaboration & LHCb Collaboration Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data. *Nature* **522**, 68–72 (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. Preprint at https://arxiv.org/abs/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. *Nat. Commun.* **5**, 4308 (2014).
    This paper introduced deep learning to high-energy physics and explains 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. *J. High Energy Phys.* **7**, 69 (2016).
    This paper started investigations of deep-learning approaches to jets in quantum chromodynamics and includes a detailed discussion of CNNs and supporting exploration of network behaviour.
51. Racah, E. et al. Revealing fundamental physics from the Daya Bay Neutrino Experiment using deep neural networks. In *15th IEEE International Conference on Machine Learning and Applications* 892–897 (IEEE, 2016).
52. Aurisano, A. et al. A convolutional neural network neutrino event classifier. *J. Instrum.* **11**, P09001 (2016).
    The paper presents the first CNN to be used for a physics analysis[70] and includes a detailed discussion of the method and comparison to more traditional neutrino-identification methods.
53. Sadowski, P, et al. Efficient antihydrogen detection in antimatter physics by deep learning. *J. Phys. Commun.* **1**, 025001 (2017).
54. Renner, J. Background rejection in NEXT using deep neural networks. *J. Instrum.* **12**, T01004 (2017).
55. Wielgosz, M., Skoczeń, A. & Mertik, M. Using LSTM recurrent neural networks for monitoring the LHC superconducting magnets. *Nucl. Instrum. Methods A* **867**, 40–50 (2017).
56. Edelen, A. L. et al. Neural networks for modeling and control of particle accelerators. *IEEE Trans. Nucl. Sci.* **63**, 878–897 (2016).
57. LeCun, Y. et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1**, 541–551 (1989).
58. Gers, F. A., Schmidhuber, J. & Cummins, F. Learning to forget: continual prediction with LSTM. *Neural Comput.* **12**, 2451–2471 (1999).
59. Cogan, J., Kagan, M., Strauss, E. & Schwartzman, A. Jet-images: computer vision inspired techniques for jet tagging. *J. High Energy Phys.* **2**, 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. D* **93**, 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. D* **95**, 014018 (2017).
62. Komiske, P. T., Metodiev, E. M. & Schwartz, M. D. Deep learning in color: towards automated quark/gluon jet discrimination. *J. High Energy Phys.* **1**, 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**, 4 (2017).
64. Kasieczka, G., Plehn, T., Russell, M. & Schell, T. Deep-learning top taggers or the end of QCD? *J. High Energy Phys.* **5**, 6 (2017).
65. Shimmin, C. et al. 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*. Report No. ATL-PHYS-PUB-2017-017, https://cds.cern.ch/record/2275641 (CERN, 2017).
67. CMS Collaboration. *New Developments for Jet Substructure Reconstruction in CMS*. Report No. CMS-DP-2017-027, https://cds.cern.ch/record/2275226 (CERN, 2017).
68. NOvA Collaboration. *The NOvA Technical Design Report*. Report No. FERMILAB-DESIGN-2007-01 (FNAL, 2007).
69. Szegedy, C. et al. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition* 1–9 (IEEE, 2015).
70. Adamson, P. et al. Constraints on oscillation parameters from v_e appearance and v_μ disappearance in NOvA. *Phys. Rev. Lett.* **118**, 231801 (2017).
71. Acciarri, R. et al. Design and construction of the MicroBooNE detector. *J. Instrum.* **12**, P02017 (2017).
72. Adamson, P. et al. Search for active-sterile neutrino mixing using neutral-current interactions in NOvA. *Phys. Rev. D* **96**, 072006 (2017).

73. Acciarri, R. et al. Convolutional neural networks applied to neutrino events in a liquid argon time projection chamber. *J. Instrum.* **12**, P03011 (2017).
74. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 061137 (2017).
75. ATLAS Collaboration. Performance of *b*-jet identification in the ATLAS experiment. *J. Instrum.* **11**, P04008 (2016).
76. CMS Collaboration. *CMS Phase 1 Heavy Flavour Identification Performance and Developments*. Report No. CERN-CMS-DP-2017-013, https://cds.cern.ch/record/2263802 (CERN, 2017).
77. ATLAS Collaboration. *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*. Report No. ATL-PHYS-PUB-2017-003, https://cds.cern.ch/record/2255226 (CERN, 2017).
78. ATLAS Collaboration. *Optimisation and Performance Studies of the ATLAS b-Tagging Algorithms for the 2017-18 LHC Run*. Report No. ATL-PHYS-PUB-2017-013, https://cds.cern.ch/record/2273281 (CERN, 2017).
79. CMS Collaboration. *Heavy Flavor Identification at CMS with Deep Neural Networks*. Report No. CMS-DP-2017-005, https://cds.cern.ch/record/2255736 (CERN, 2017).
80. Guest, D. et al. Jet flavor classification in high-energy physics with deep neural networks. *Phys. Rev. D* **94**, 112002 (2016).
81. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
82. ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data. *Eur. Phys. J. C* **77**, 195 (2017).
83. The CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *J. Instrum.* **10**, P06005 (2015).
84. Sachdev, K. *Muon Neutrino to Electron Neutrino Oscillation in NOvA*. PhD thesis, Univ. Minnesota (2015).
85. Chatrchyan, S. et al. Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons. *J. High Energy Phys.* **5**, 104 (2014).
86. van der Maaten, L. Accelerating t-SNE using tree- based algorithms. *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).
87. Louppe, G., Cho, K., Becot, C. & Cranmer, K. QCD-aware recursive neural networks for jet physics. Preprint at https://arxiv.org/abs/1702.00748 (2017).
88. Goodfellow, I. J. et al. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **27**, 2672–2680 (2014).
89. Rezende, D.J., Mohamed, S. & Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. *J. Mach. Learn. Res. Worksh. Conf. Proc.* **32**, 1278–1286 (2014).
90. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. Preprint at https://arxiv.org/abs/1312.6114 (2014).
91. Paganini, M., de Oliveira, L. & Nachman, B. Accelerating science with generative adversarial networks: an application to 3D particle showers in multilayer calorimeters. *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. In *NIPS Deep Learning for Physical Sciences Workshop* (NIPS, 2017).
93. Louppe, G., Kagan, M. & Cranmer, K. Learning to pivot with adversarial networks. *Adv. Neural Inf. Process. Syst.* **30**, 981–990 (2017).
    This paper forms part of a collection of works that present a more nuanced loss function and, along with similar work for BDTs (see section 'Conclusions and outlook'), could lead to a new paradigm for training machine-learning models in high-energy physics.
94. Stevens, J. & Williams, M. uBoost: a boosting method for producing uniform selection efficiencies from multivariate classifiers. *J. Instrum.* **8**, P12013 (2013).
95. Rogozhnikov, A., Bukva, A., Gligorov, V. V., Ustyuzhanin, A. & Williams, M. New approaches for boosting to uniformity. *J. Instrum.* **10**, T03002 (2015).
96. Dery, L. M., Nachman, B., Rubbo, F. & Schwartzman, A. Weakly supervised classification in high energy physics. *J. High Energy Phys.* **5**, 145 (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 $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decays. *Phys. Rev. Lett.* **115**, 161802 (2015).
99. The ATLAS collaboration. Search for the $b\bar{b}$ decay of the standard model Higgs boson in associated (W/Z)H production with the ATLAS detector. *J. High Energy Phys.* **1**, 69 (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. D* **89**, 012003 (2014).
101. CMS Collaboration. *2015 ECAL Detector Performance Plots*. Report No. CMS-DP-2015-057, https://cds.cern.ch/record/2114735 (CERN, 2015).