PROCEEDINGS OF SCIENCE

Novel machine learning applications at the LHC

Javier M. Duarte* **on behalf of the ALICE, ATLAS, CMS, and LHCb Collaborations**

Department of Physics
University of California San Diego
9500 Gilman Drive
La Jolla, CA 92093

E-mail: jduarte@ucsd.edu

Machine learning (ML) is a rapidly growing area of research in the field of particle physics, with a vast array of applications at the CERN LHC. ML has changed the way particle physicists conduct searches and measurements as a versatile tool used to improve existing approaches and enable fundamentally new ones. In these proceedings, we describe novel ML techniques and recent results for improved classification, fast simulation, unfolding, and anomaly detection in LHC experiments.

42nd International Conference on High Energy Physics (ICHEP2024)
18-24 July 2024
Prague, Czech Republic

*Speaker

PoS(ICHEP2024)012

1. Introduction

Particle physicists have a long history of developing and applying machine learning (ML) techniques. From early applications of neural networks to charged particle tracking in the 1980s [1] to the Higgs boson discovery in 2012, in which boosted decision trees improved the sensitivity to the $H \to \tau\tau$ decay mode [2], ML has changed the way particle physicists conduct searches and measurements. It is an essential and versatile tool that we use to improve existing approaches, and it enables fundamentally new approaches. In recent years, the subfield of ML in particle physics has grown exponentially in the number of publications and expanded to cover a wide variety of topics and use cases, as indexed by the HEP ML Living Review [3].

In these proceedings, we present selected recent results that highlight how LHC experiments are applying novel ML techniques. In particular, we briefly describe the ML techniques and results for improved classification, faster simulation, unfolding, and anomaly detection.

2. Improved classification

Great strides have been made to leverage rich low-level information for a variety of tasks, including jet classification, by considering different *representations*. Traditionally, jets are pre-processed into a vector of high-level features for use with ML methods like feedforward neural networks or boosted decision trees. More recently, jets have been represented as sequences, images, or graphs; the latter can be processed by graph neural networks [4] or attention-based transformers [5]. For example, major improvements have been demonstrated in CMS with ParticleNet [6] and Particle Transformer (ParT) [7] and in ATLAS with the GN1 [8] and GN2 [9] heavy flavor taggers. The improvement over the years of increasingly sophisticated b-tagging algorithms in ATLAS is shown in Fig. 1 (left). In particular, the latest GN2 algorithm boasts a factor of 4.2 improvement in light jet rejection over the baseline. Figure 1 (right) also shows the power of the CMS ParticleNet $X(qq)$ large-radius jet tagger in a search for light Z' resonances [10]. The tagger is able to reduce the substantial QCD multijet background to reveal the prominent $W(qq)$ resonance in the jet soft drop mass spectrum.

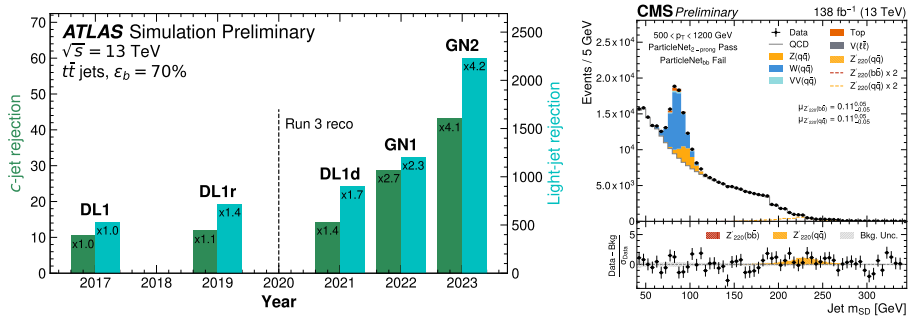

Figure 1: The c and light jet rejection of the different ATLAS flavor tagging algorithms over time in Monte Carlo simulation (left) [9]. The jet soft drop mass distribution after two-prong and light-flavor ParticleNet tagger selections, illustrating the prominent $W(qq)$ resonance (right) [10].

Building on the success of ParticleNet in CMS for large-radius jet tagging, which focused on discriminating $X(qq)$, $X(cc)$, $X(bb)$, and QCD jets, a new algorithm, dubbed Global ParT

(GloParT) [11] is trained on a much larger set of classes, including all-hadronic and semileptonic $X(VV)$ decays, as shown in Fig. 2 (left). This algorithm is based on ParT [7], which leverages a learned "attention" and pairwise features to give more weight to certain particles, and disregard others, in order to infer the origin of jets. Since this algorithm is trained in Monte Carlo (MC) simulation, it can be a challenge to calibrate in data when there is no SM analogue for the signal, such as the $H \to VV \to 4q$, that we can isolate. However, a new calibration technique uses the Lund jet plane [12]. Effectively, this technique measures data-to-simulation corrections per quark subjet, allowing the appropriate corrections to be determined with a readily available data sample of $W(qq)$ jets.

This tagger enables a new search for nonresonant, boosted $HH \to b\bar{b}VV \to b\bar{b}4q$ [11]. The vector boson fusion (VBF) HH production mode is especially sensitive to a coupling modifier of the two-Higgs-boson-two-vector-boson ($HHVV$) interaction in the SM, known as κ_{2V}. If κ_{2V} differs from the SM value of 1, the differential VBF HH production cross section dramatically increases at high m_{HH}, resulting in enhanced HH production. Figure 2 (center) displays the $b\bar{b}$-candidate jet mass after a tight GloParT $H \to 4q$ tagger and VBF-like selection, where the BSM signal would be clearly visible. The search provides the second-best constraint on this coupling in the CMS experiment, as illustrated in Fig. 2 (right).

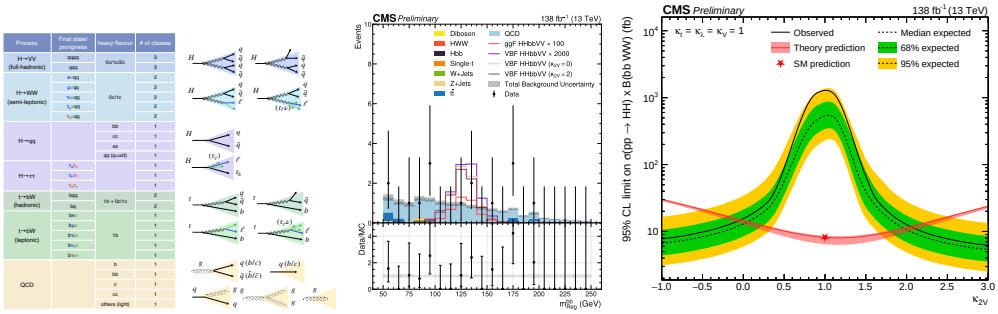

Figure 2: Full set of training jet classes for GloParT (left). Distributions of the $b\bar{b}$-candidate jet mass in the VBF signal region (center). Upper limits on the inclusive HH production cross section as a function of κ_{2V} (right).

For small-radius jets in CMS, a similar algorithm has been developed that unifies heavy flavor tagging, hadronic tau tagging, jet energy regression, and jet energy resolution estimation, called unified particle transformer (UParT) [13]. An additional novel aspect is the use of a rectified normed gradient method (R-NGM) adversarial training to improve model robustness. In this strategy, the inputs are perturbed by adding a term proportional to the absolute value of the gradient of the loss function, which maximally disrupts the performance. The trained model can then generalize better when faced with perturbed or mismodeled inputs. This is illustrated in the receiver operating characteristic (ROC) curves shown in Fig. 3 (left), where UParT trained with R-NGM shows a substantial improvement compared to the nominal training when evaluated with the perturbed inputs, while preserving a similar performance when evaluated with the nominal inputs. Figure 3 (right) also demonstrates the superior jet energy regression performance of UParT over a previous algorithm based on ParticleNet.

In the ALICE experiment, transformers have been used for particle identification, which results

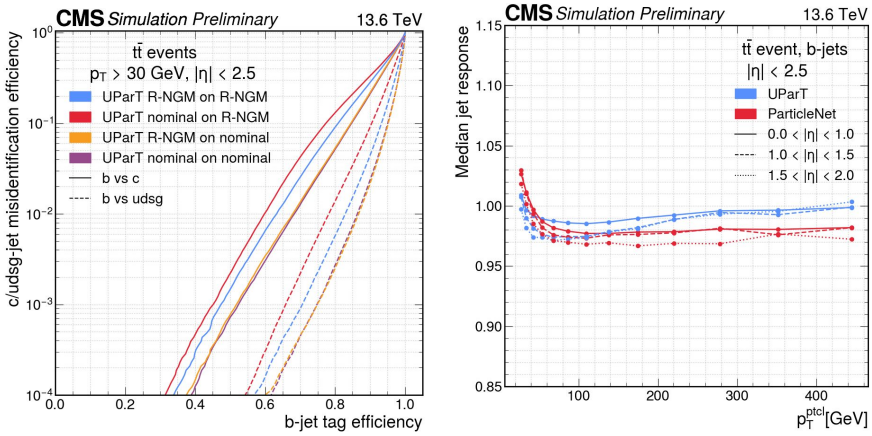

Figure 3: The ROC curves for UParT trained with R-NGM or nominal samples and evaluated with R-NGM or nominal samples (left). Median of the raw regressed jet energy response for UParT and ParticleNet (right).

in higher purity and efficiency than standard methods, even for data with missing values due to limited detector efficiency and acceptance [14, 15]. The proposed model architecture is based on a transformer, as shown in Fig. 4 (left). A domain-adversarial neural network approach consisting of three neural networks is also proposed to mitigate differences between the two "domains" of data and simulation. In this setup, a featurizer maps the original input to domain-invariant features that are provided to the particle classifier and the domain classifier, which enforces the domain invariance of the extracted features through an adversarial training, as shown in Fig. 4 (right).

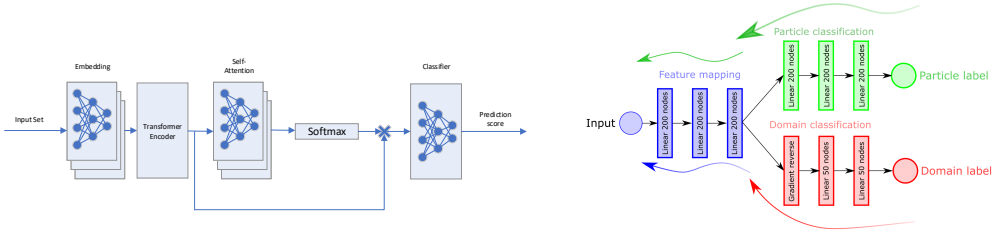

Figure 4: Transformer architecture for particle identification in ALICE (left). Layered blocks are applied separately to each vector in a set. Single blocks are applied to their input as a whole. Domain-adversarial neural network training setup (right).

Finally, a new systematic-aware neural network training (SANNT) has been developed in the context of a $H \to \tau\tau$ search in CMS [16]. Traditionally, classifier neural networks are trained using the cross entropy loss function (cross entropy neural network training, or CENNT), which optimizes for signal versus background discrimination without considering systematic effects that may influence the ultimate figure of merit: the measurement uncertainty Δr_s on a physics model's parameter of interest r_s, such as a signal strength. Alternatively, by implementing the full analysis chain, including systematic uncertainties, in a differentiable way, it is possible to directly minimize δr_s as the neural network loss function using gradient descent. Flow charts of these two contrasting approaches are shown in Fig. 5 (left). A crucial ingredient is the choice of custom functions $\mathcal{B}_i$ to substitute for the gradients of the non-differentiable histogram operation $\hat{\mathbf{y}} \to H(\hat{\mathbf{y}})$. Based on

SANNT (CENNT), values of $\Delta r_s = ^{+0.47}_{-0.44}$ ($\Delta r_s = ^{+0.62}_{-0.60}$) for the $H \to \tau\tau$ signal strength are obtained, corresponding to a 25% reduction in the measurement uncertainty for SANNT compared to CENNT as shown in Fig. 5 (right). For the first time in an analysis of this complexity, the significant gains of SANNT over CENNT are demonstrated.

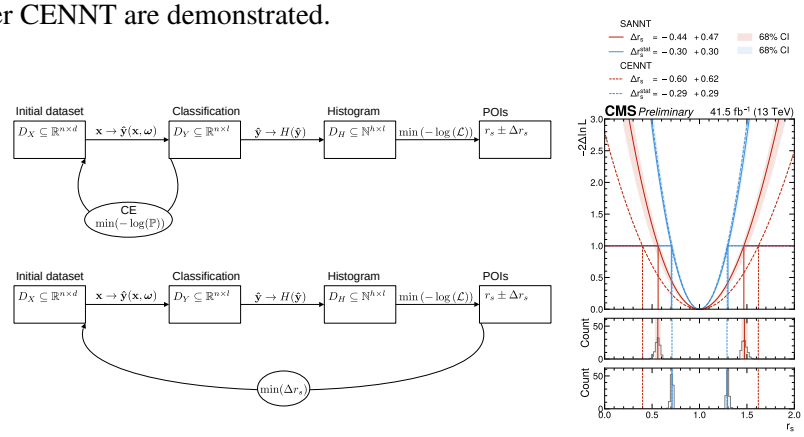

Figure 5: Flow chart of a CENNT (upper left) and SANNT (lower left). Negative log of the profile likelihood $-2\Delta \ln L$ as a function of r_s, taking into account (red) all and (blue) only the statistical uncertainties in Δr_s (right). The results as obtained from CENNT (SANNT) are indicated by the dashed (solid) lines.

3. Fast simulation

Generating large-scale, realistic simulated data samples is a key driver of the increased CPU needs for the high-luminosity LHC. Machine learning methods can be used to "short cut" traditional simulation, producing simulated data samples of sufficient quality for analysis at a fraction of the CPU cost. The ATLAS Collaboration has investigated the use of variational autoencoders (VAEs) and generative adversarial networks (GANs) to model the response of the ATLAS electromagnetic calorimeter to photons of various energies [17]. Compared to a full detector simulation using GEANT4, both VAEs and GANs are found to be capable of quickly simulating electromagnetic showers with correct total energies (Fig. 6, left and center) and stochasticity, though the modeling of some shower shape distributions requires more refinement.

Another approach to reduce the computational costs of simulation is the use of a neural network to reweight from one simulated sample to another one with different physics model parameters or to match more accurate (higher order in perturbation theory) calculations. This method circumvents the need to simulate the detector response multiple times by reusing a single sample with different learned event weights. The method relies on the *likelihood ratio trick*, in which a classifier $f(x)$ trained to distinguish samples from two different probability distribution functions, $p_0(x)$ or $p_1(x)$, approximates the likelihood ratio $f(x)/(1-f(x)) \approx p_0(x)/p_1(x)$, which can be used as an event weight [18]. The CMS Collaboration has used this method for several applications in top quark physics, including to reweight MC simulation to account for higher-order theory predictions, as shown in Fig. 6 (right) [19].

4. Unfolding

Machine learning methods can also be used to "unfold" detector effects and measure the differential distributions of physics observables at the particle level. The ATLAS Collaboration

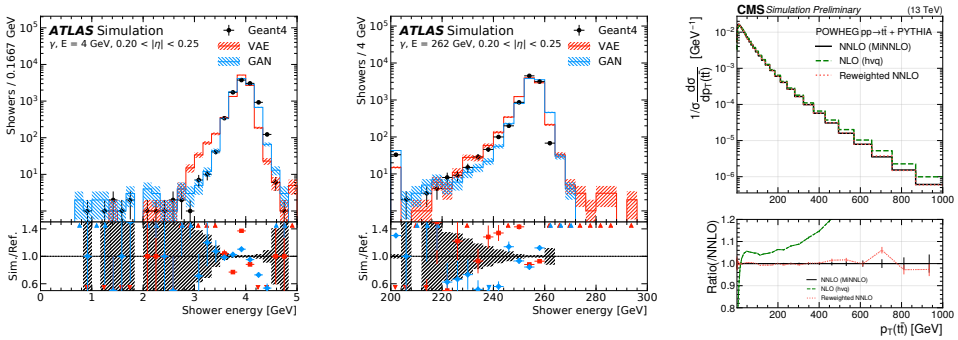

Figure 6: Total energy response of the calorimeter to photons with an energy of 4 GeV (left) and 262 GeV (center). The calorimeter response for the GEANT4 training data (black markers) are shown as reference points and compared with those from a VAE (solid red line) and a GAN (solid blue line). Distributions in p_{T} of the $t\bar{t}$ system obtained from simulations at next-to-next-to-leading-order (NNLO) accuracy (black solid lines), next-to-leading-order (NLO) accuracy (green dashed lines), and NLO reweighted to NNLO (red dotted lines) (right).

has employed the OmniFold method [20] to produce a simultaneous measurement of 24 Z+jets observables [21]. In this method, the detector-level MC simulation is first corrected by a learned weighting function $\omega(\vec{x}_r)$ to match data. Then, the particle-level MC simulation is corrected by another learned weighting function $\nu(\vec{x}_p)$ to match the $\omega(\vec{x}_r)$-weighted MC simulation. The method is iterated four more times, to achieve $\nu(\vec{x}_p)$-weighted MC events whose event yields and kinematics match those observed in data. Unlike previous fiducial differential cross section measurements, the result is presented as an unbinned dataset of weighted particle-level events [22], allowing for new observables to be constructed from the 24 measured observables as shown in Fig. 7 (lower right).

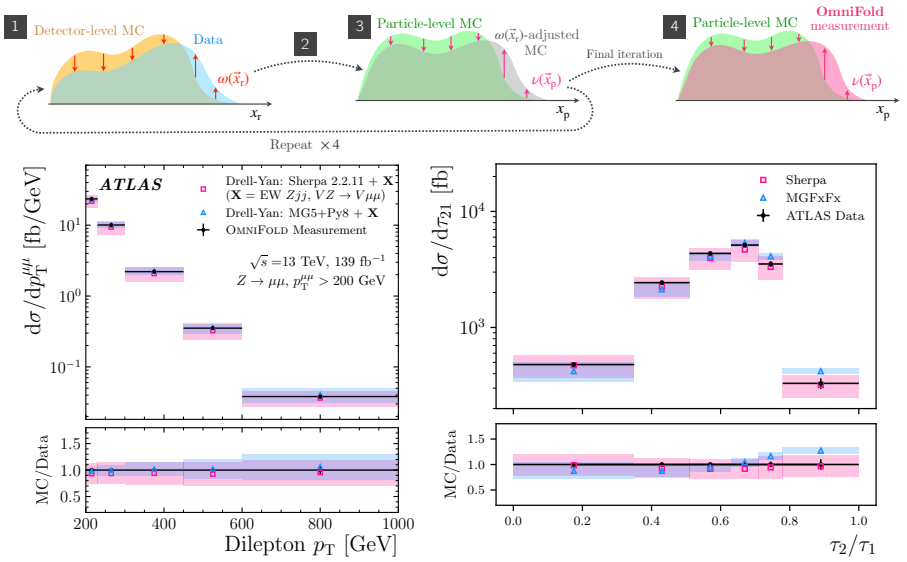

Figure 7: Illustration of the OmniFold method (upper). Measured differential cross sections compared with particle-level predictions from Sherpa and MadGraph for one of the 24 directly measured observables in Ref. [21], the dilepton p_{T} (lower left), and another observable that can be derived with the released dataset the leading jet substructure variable τ_2/τ_1 (lower right).

5. Anomaly detection

If the form of new physics is not known, it can be difficult if not impossible to design a search to target it. Anomaly detection methods offer the promise of sensitivity to a broad range of new physics signatures, enabling model-agnostic searches. There are several varieties of ML-based anomaly detection techniques, ranging from unsupervised methods, in which no labels are known during training, to weakly supervised methods, in which labels are partially known. A classic example of an unsupervised method is a (variational) autoencoder, or (V)AE, that compresses input data then attempts to reconstruct it. An anomaly score can then be built based on the distance between the input and the output—the assumption being that those examples that are misreconstructed are outside of the training domain and thus anomalous.

In the CMS experiment, five different anomaly detection techniques are used to search for new physics in dijet events [23], including a VAE with quantile regression (VAE-QR) to ensure a constant efficiency as a function of the invariant mass of the dijet. Figure 8 (left) shows the dijet invariant mass spectrum and resulting background fit to the data for the VAE-QR method, along with example signals. The methods reduce the cross section needed for a 5σ discovery compared to the inclusive dijet search by up to factor of 7, as shown in Fig. 8 (right).

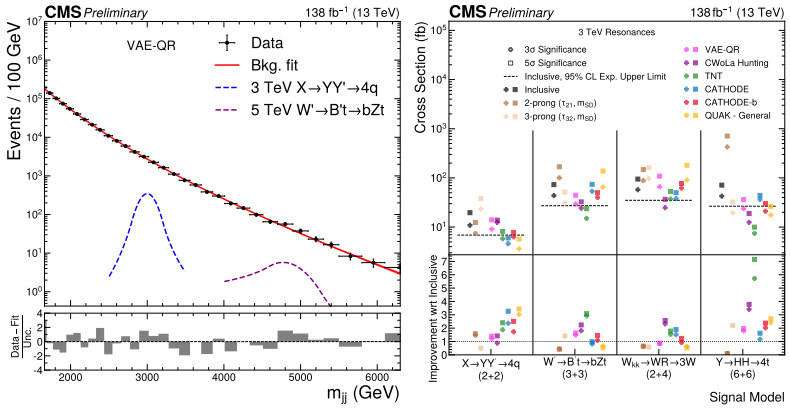

Figure 8: The dijet invariant mass spectrum and resulting background fit to the data for the VAE-QR method, along with example signals (left). The cross section that would lead to an expected 3σ (5σ) excess is shown as a cross (square) marker for the six anomaly detection methods (six colors), an inclusive dijet search (black), and traditional substructure cuts targeting two-pronged (dark brown) or three-pronged decays (tan) (right).

Anomaly detection algorithms can be applied to select events directly at the trigger level. Such an algorithm, called AXOL1TL, has been developed for the CMS level-1 global trigger, consisting of a VAE trained on zero bias data events using the kinematic information of up to 10 jets, 4 muons, 4 electrons, and the missing transverse momentum in each event as input [24, 25]. The encoder is compiled into firmware for an FPGA with the hls4ml package [26], and an anomaly score is calculated based on the latent representation. Figure 9 (left) shows the score distributions during 2024 data collection. In particular, the pure contribution shows that AXOLT1TL selects unique events relative to existing level-1 trigger. However, there is also a preference for high-multiplicity events (Fig. 9, right), which leads to an undesirable pileup dependence.

At times, autoencoders can be too good at reconstructing signal, meaning that signal is not flagged as anomalous due to a low reconstruction error. The normalized autoencoder approach [27]

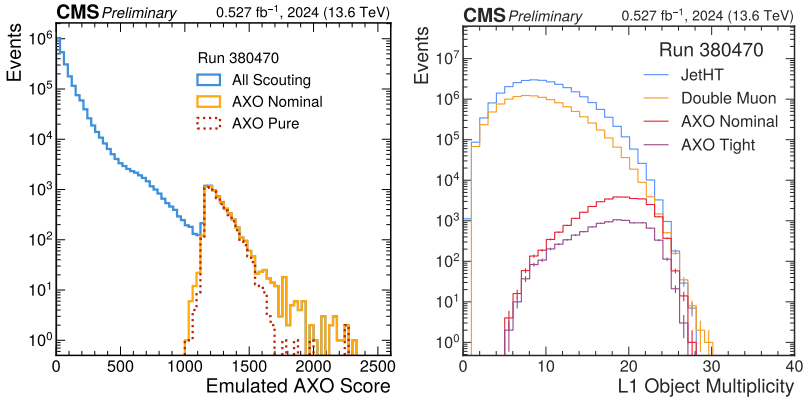

Figure 9: AXOL1TL score distributions for the HLT scouting partial dataset for CMS Run 380470 on May 7, 2024. Scores are shown for all events triggered by the nominal seed, and those only triggered by the AXOL1TL triggers and no other level-1 triggers (pure). Level-1 trigger object multiplicity distributions for jet H_{T}, double muon, and AXOL1TL triggers (right).

is designed to mitigate this issue by aligning the low reconstruction error phase space with the background phase space. The LHCb experiment has implemented this approach to trigger on long-lived particles decaying in the muon chambers [28]. The idea is to promote the autoencoder to an energy-based model that models the data as $p_\theta(x) \propto \exp\left[\mathrm{MSE}(x, \mathrm{AE}(x))\right]$; the parameters of the autoencoder are then obtained by minimizing the negative log likelihood of the data $-\ln p_\theta(x)$. In practice, this is approximately minimized through a clever reframing using Markov chain Monte Carlo to sample from $p_\theta(x)$. With this change in training, the normalized autoencoder can capture more anomalous signals, as shown in Fig. 10

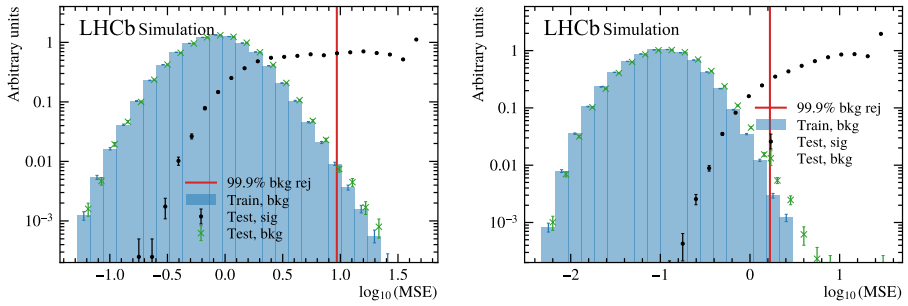

Figure 10: Distribution of the predictions for the background and signal samples, using the autoencoder (left) and the normalized autocndoder (right). The signal consists of long-lived axions produced in a Higgs boson decay $H \to AA$, decaying to tau leptons $A \to \tau\tau$, where each tau lepton decays to three pions.

6. Summary and outlook

There is a dizzying array of ML opportunities, innovations, and applications in LHC experiments, which directly impact physics results. Physicists are concerned not only with performance but also robustness, interpretability, and insensitivity to modeling uncertainties. The use cases of ML extend beyond classification to include simulation, unfolding, anomaly detection, and more. ML methods have enabled new searches and measurements that were previously impossible.

References

[1] B. H. Denby, "Neural Networks and Cellular Automata in Experimental High-energy Physics", *Comput. Phys. Commun.* **49** (1988) 429, doi:10.1016/0010-4655(88)90004-5.

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

[3] HEP ML Community, "A Living Review of Machine Learning for Particle Physics", 2021. https://iml-wg.github.io/HEPML-LivingReview.

[4] E. A. Moreno et al., "JEDI-net: a jet identification algorithm based on interaction networks", *Eur. Phys. J. C* **80** (2020) 58, doi:10.1140/epjc/s10052-020-7608-4, arXiv:1908.05318.

[5] A. Vaswani et al., "Attention is all you need", in *Advances in Neural Information Processing Systems 30*, I. Guyon et al., eds., volume 30, p. 5998. Curran Associates, Inc., 2017. arXiv:1706.03762.

[6] H. Qu and L. Gouskos, "ParticleNet: Jet tagging via particle clouds", *Phys. Rev. D* **101** (2020) 056019, doi:10.1103/PhysRevD.101.056019, arXiv:1902.08570.

[7] H. Qu, C. Li, and S. Qian, "Particle Transformer for jet tagging", in *Proceedings of the 39th International Conference on Machine Learning*, p. 18281. 2022. arXiv:2202.03772.

[8] ATLAS Collaboration, "Graph Neural Network Jet Flavour Tagging with the ATLAS Detector", ATLAS Public Note ATL-PHYS-PUB-2022-027, CERN, 2022.

[9] ATLAS Collaboration, "Jet Flavour Tagging With GN1 and DL1d", ATLAS Public Plots ATLAS-PLOTS-FTAG-2023-01, 2023.

[10] CMS Collaboration, "Search for low mass vector and scalar resonances decaying into quark-antiquark pairs", CMS Physics Analysis Summary CMS-PAS-EXO-24-007, 2024.

[11] CMS Collaboration, "Search for highly energetic double Higgs boson production in the two bottom quark and two vector boson all-hadronic final state", CMS Physics Analysis Summary CMS-PAS-HIG-23-012, 2024.

[12] CMS Collaboration, "Lund Plane Reweighting for Jet Substructure Correction", CMS Detector Performance Note CMS-DP-2023-046, 2023.

[13] CMS Collaboration, "A unified approach for jet tagging in Run 3 at $\sqrt{s}$ = 13.6 TeV in CMS", CMS Detector Performance Note CMS-DP-2024-066, 2024.

[14] M. Kasak et al., "Machine-learning-based particle identification with missing data", *Eur. Phys. J. C* **84** (2024) 691, doi:10.1140/epjc/s10052-024-13047-3, arXiv:2401.01905.

[15] ALICE Collaboration, "Particle identification with machine learning from incomplete data in the ALICE experiment", *JINST* **19** (2024) C07013, doi:10.1088/1748-0221/19/07/C07013, arXiv:2403.17436.

[16] CMS Collaboration, "Development of systematic-aware neural network trainings for binned-likelihood-analyses at the LHC", CMS Physics Analysis Summary CMS-PAS-MLG-23-005, 2024.

[17] ATLAS Collaboration, "Deep Generative Models for Fast Photon Shower Simulation in ATLAS", *Comput. Softw. Big Sci.* **8** (2024) 7, doi:10.1007/s41781-023-00106-9, arXiv:2210.06204.

[18] A. Andreassen and B. Nachman, "Neural Networks for Full Phase-space Reweighting and Parameter Tuning", *Phys. Rev. D* **101** (2020) 091901, doi:10.1103/PhysRevD.101.091901, arXiv:1907.08209.

[19] CMS Collaboration, "Reweighting of simulated events using machine learning techniques in CMS", CMS Physics Analysis Summary CMS-PAS-MLG-24-001, 2024.

[20] A. Andreassen et al., "OmniFold: A Method to Simultaneously Unfold All Observables", *Phys. Rev. Lett.* **124** (2020) 182001, doi:10.1103/PhysRevLett.124.182001, arXiv:1911.09107.

[21] ATLAS Collaboration, "A simultaneous unbinned differential cross section measurement of twenty-four Z+jets kinematic observables with the ATLAS detector", 2024. arXiv:2405.20041. Submitted to *Phys. Rev. Lett.*

[22] ATLAS Collaboration, "ATLAS OmniFold 24-Dimensional Z+jets Open Data", 2024. doi:10.5281/zenodo.11507450.

[23] CMS Collaboration, "Model-agnostic search for dijet resonances with anomalous jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-EXO-22-026, 2024.

[24] CMS Collaboration, "Anomaly Detection in the CMS Global Trigger Test Crate for Run 3", CMS Detector Performance Note CMS-DP-2023-079, 2023.

[25] CMS Collaboration, "2024 Data Collected with AXOL1TL Anomaly Detection at the CMS Level-1 Trigger", CMS Detector Performance Note CMS-DP-2024-059, 2024.

[26] J. Duarte et al., "Fast inference of deep neural networks in FPGAs for particle physics", *JINST* **13** (2018) P07027, doi:10.1088/1748-0221/13/07/P07027, arXiv:1804.06913.

[27] B. M. Dillon et al., "A normalized autoencoder for LHC triggers", *SciPost Phys. Core* **6** (2023) 074, doi:10.21468/SciPostPhysCore.6.4.074, arXiv:2206.14225.

[28] LHCb Collaboration, "Anomaly detection for a new trigger line for showers in the muon detectors", LHCb Figures LHCB-FIGURE-2024-015, 2024.