Eur. Phys. J. C (2024) 84:442
https://doi.org/10.1140/epjc/s10052-024-12790-x

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

Machine learning classification of sphalerons and black holes at the LHC

Aurora Singstad Grefsrud[1], **Trygve Buanes**[1], **Fotis Koutroulis**[3], **Anna Lipniacka**[2], **Rafał Maselek**[3,5], **Andreas Papaefstathiou**[4], **Kazuki Sakurai**[3,a], **Therese B. Sjursen**[1], **Igor Slazyk**[1]

[1] Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Postbox 7030, 5020 Bergen, Norway
[2] Department of Physics and Technology, University of Bergen, Postboks 7803, 5020 Bergen, Norway
[3] Institute of Theoretical Physics, Faculty of Physics, University of Warsaw, ul. Pasteura 5, 02-093 Warsaw, Poland
[4] Department of Physics, Kennesaw State University, 830 Polytechnic Lane, Marietta, GA 30060, USA
[5] Laboratoire de Physique Subatomique et de Cosmologie (LPSC), Université Grenoble-Alpes, CNRS/IN2P3, 53 Avenue des Martyrs, 38026 Grenoble, France

Received: 19 January 2024 / Accepted: 11 April 2024

Abstract In models with large extra dimensions, "miniature" black holes (BHs) might be produced in high-energy proton–proton collisions at the Large Hadron Collider (LHC). In the semi-classical regime, those BHs thermally decay, giving rise to large-multiplicity final states with jets and leptons. On the other hand, similar final states are also expected in the production of electroweak sphaleron/instanton-induced processes. We investigate whether one can discriminate these scenarios when BH or sphaleron-like events are observed in the LHC using machine learning (ML) methods. Classification among several BH scenarios with different numbers of extra dimensions and the minimal BH masses is also examined. In this study we consider three ML models: XGBoost algorithms with (1) high- and (2) low-level inputs, and (3) a Residual Convolutional Neural Network. In the latter case, the low-level detector information is converted into an input format of three-layer binned event images, where the value of each bin corresponds to the energy deposited in various detector subsystems. We demonstrate that only a small number of detected events are sufficient to effectively discriminate between the sphaleron and BH processes. Separation between BH scenarios with different minimal masses is possible with an order of 10 events passing the preselection. A sufficient number of events could be observed in combined Run-2 and -3 data, if the production cross section is not much smaller than the present limit ~ 0.1 fb. We find, however, that a large number of events is needed to discriminate between BH hypotheses with the same minimal BH mass, but different numbers of extra dimensions.

[a] e-mail: kazuki.sakurai@fuw.edu.pl (corresponding author)

1 Introduction

The Standard Model (SM) of particle physics has been extremely successful in describing particle interactions below the TeV scale. However, the SM has several theoretical issues, such as the hierarchy problem, as well as phenomenological problems like the inability to account for the existence of dark matter and the observed asymmetry between matter and antimatter in the Universe. These problems suggest that a more fundamental theory underlying the SM must exist, and may emerge at the energy scale around, or higher, than the TeV scale.

An attractive solution to the hierarchy problem is a scenario with *Large Extra Dimensions* (LEDs) [1–3]. This model postulates that all matter and gauge fields, except for gravity, are restricted to live on a $(3+1)$-dimensional hypersurface, called a 3-brane, which is embedded in the higher dimensional spacetime. Assuming all extra dimensions orthogonal to the 3-brane are compactified, the traditional Planck scale $M_P \sim 10^{18}$ GeV is understood as an effective scale, derived from the fundamental higher-dimensional Planck scale M_* with the relation:

$$M_*^{2+n} \sim \frac{M_P^2}{R^n}, \tag{1.1}$$

where we have assumed the existence of n extra dimensions with the common size R. If the size of the extra dimensions is much larger than the traditional Planck length, $R \gg 1/M_P$, the fundamental Planck scale M_* can be much smaller than the usual Planck scale M_P. In particular, the hierarchy problem is solved if M_* is around the TeV scale [1–3].

Published online: 30 April 2024

This scenario has a striking implication for new physics searches at high-energy colliders, such as the Large Hadron Collider (LHC) at CERN. For the fundamental Planck scale around or smaller than the TeV scale, $M_* \lesssim \mathcal{O}(1)$ TeV, high-energy colliders can offer trans-Planckian particle collisions with $\sqrt{\hat{s}} > M_*$. In this situation, the gravitational interaction becomes very strong and the colliding particles may collapse into a black hole (BH) [4–11]. In the semi-classical regime, where the BH mass is significantly larger than M_*, such a "miniature" BH, once formed, will quickly "evaporate" through Hawking radiation [12].[1] The spectrum of emitted particles can be understood as the thermal radiation characterized by the Hawking temperature T_H of the BH at a given stage of the evaporation [13,18]. Those outgoing particles, however, must travel through a strong gravitational potential created by the BH, and the spectrum is distorted from the blackbody profile for an observer located infinitely far from the BH [19,20]. The deviation from the blackbody spectrum is described by the so-called "greybody factor", which carries the information of extra dimensions [13].

The last observation poses an interesting question to collider physics: "can we discriminate one extra dimension scenario from another with the BH signature at a collider, if observed?", or more concretely, for example, "can we identify the number of extra dimensions by analysing BH events?". Since the greybody factor depends on the parameters of extra dimensions in a subtle way, answering these questions is non-trivial.

There even exists a similar collider signature that has a completely different origin, due to electroweak (EW) sphaleron/instanton-induced processes [21–24]. These are non-perturbative processes within the SM, which go over (or penetrate via quantum tunneling) a potential barrier separating two distinctive EW vacua: the Higgs and EW gauge field configurations in these vacua are characterised by different values of the topological winding number, $N_{\rm CS}$ (the Chern–Simons number). The EW sphaleron plays a crucial role in many proposed scenarios of baryogenesis (see, e.g., Ref. [25] for a review). The theoretical prediction for the production cross section of these non-perturbative processes at high-energy colliders suffers from large uncertainties, and whether such processes are observable in the foreseeable future is still under debate [26–45]. Although the cross section is largely unknown, the final state of the process is anticipated to possess certain characteristic features. Firstly, the minimum height of the potential barrier, $E_{\rm sph}$, is known to good accuracy, $E_{\rm sph} \simeq 9.1$ TeV [21,22]. Therefore, one expects that those processes can occur only at very high energies, at least with $\sqrt{\hat{s}} \gtrsim E_{\rm sph}$. Secondly, since the anomaly connects the change of $N_{\rm CS}$, and the change of the fermion number that couples to the $SU(2)_L$ gauge bosons, *all* left-handed fermions, (3 quarks + 1 lepton) × (3 generations), of the SM must be involved in the interaction [46,47]. Therefore, 'consuming' two light-flavor quarks in the initial state, the final state must contain at least seven anti-quarks and three anti-leptons (plus some EW bosons). Such a high-energy and high-multiplicity final state resembles the BH events mentioned. It is therefore a non-trivial task to discriminate the EW sphaleron/instanton-induced processes from the miniature BH signature in the LED scenario.

A traditional way of tackling the above questions is to analyse various event variables built out of reconstructed objects. In particular, in Ref. [48], the CMS collaboration has performed an analysis to search for semi-classical BHs and EW sphalerons, resulting in the cross section limits of the order of 0.1 fb.[2] The CMS event selection is based on the reconstructed high-p_T objects, with $p_T > 70$ GeV, including jets, isolated electrons, muons and photons. Although this approach appears to be powerful for searches, it has several drawbacks when applied to discrimination problems in high-multiplicity final states. For example, we will show later that isolated leptons are potentially a powerful discriminator between the BH and EW sphaleron scenarios. However, charged leptons arise only in some fraction of signal events. Moreover, those leptons are often rejected by the isolation criterion in the busy environment of a high-multiplicity event. As a result, the majority of signal events contain no reconstructed leptons. This is a crucial problem for our case since the expected signal cross sections are low ($\lesssim 0.1$ fb). Reference [5] proposed a method to measure the number of extra dimensions by fitting the Hawking temperature formula with the electron/photon spectrum in BH events. Given the current cross section limit, however, this method is no longer available as it requires large statistics. Ideally, the optimal discrimination method should use as much information as possible in the high multiplicity event data, rather than looking at a particular object. Finding the optimal analysis method or constructing a good discrimination variable, however, becomes

[1] As opposed to the semi-classical regime, production and decay of quantum BHs (QBHs) have been discussed [13,14], and collider searches have been carried out. Those searches are based on the assumption that QBHs decay exclusively into two high-energy SM particles. The current limit depends on the decay channels: $\lesssim 1$ fb for dijet [15] and $\lesssim 0.1$ fb for lepton+jet [16] and charged lepton flavour violating final channels (QBH $\to e\mu, e\tau, \mu\tau$) [17]. If these limits are considered seriously, there is little chance that the LHC can observe semi-classical BHs. However, the dynamics of QBHs are much less known than those of their semi-classical counterparts, and in this paper, we assume the cases where the current limit on QBH is not applicable.

[2] They set the model-independent limit $\lesssim 0.1$ fb on the cross section times acceptance in the signal regions with $N \gtrsim 5$ and $S_T \gtrsim 7$ TeV or $N \gtrsim 10$ and $S_T \gtrsim 4$ TeV (see the definitions of N and S_T in the next section). In the sphaleron scenario, this corresponds to the prefactor $\lesssim 0.02$ with the threshold energy $E_{\rm thr}^{\rm sph} \gtrsim 9$ TeV. For the BH scenario with a certain assumption (non-rotating, $n = 6$, the fundamental Planck scale $M_* = 4$ TeV) and the Hoop conjecture [49], this limit translates into the minimal BH mass $M_{\rm min} \gtrsim 10$ TeV. For more details, see Ref. [48].

increasingly difficult when the number of final state objects increases.

The identification of the number of extra dimensions, and the discrimination between the BH and EW sphaleron/instanton events, is essentially a classification problem, which machine learning (ML) methods have been proven to be exceptionally good at. Reference [50] studied a way to look for EW sphaleron events in cosmic ray showers with ML. The application of ML to collider physics has also been an active research field in the last decade. Recent development includes the application of deep learning, i.e., neural networks (NNs) to, for example, triggering, background estimation, jet tagging, and event classification (see, e.g., Ref. [51] for a comprehensive list of references). More recently, deep learning methods utilising low-level whole-event data (e.g., the energy deposits of the entire calorimeter) have been investigated [52–54]. These studies represented entire collider events as images, and processed them using Convolutional Neural Networks (CNNs) [55]. They have demonstrated that such methods can exceed the sensitivity of standard approaches with high-level inputs, such as the four-momenta of reconstructed objects. Classifying events using low-level inputs has an obvious advantage in our problem, since the object reconstruction process is largely omitted, and most detector information is kept and used in the analysis. It should be noted, however, that this approach has certain limitations in high pile-up conditions, when the detector is swamped by energy deposits that do not belong to the main hard interaction of interest. Reconstructions of jets pointing to the main vertex of interest, and rejecting the deposits not belonging to them, typically serve as pile-up mitigation. In what follows, we assume that pile-up mitigation can be performed, and we do not consider energy deposits coming from pile-up interactions. Furthermore, we assume that the SM backgrounds can be suppressed with the CMS-inspired selections [48], which are applied to our data.

In this paper, we study the discrimination of EW sphaleron and five different BH scenarios, using three different ML methods. The aim is to see if we can separate these scenarios with a reasonable number of events that can be collected in the on-going and future LHC runs. Three ML methods are examined. The first two are based on the state-of-the-art XGBoost library, which constructs a boosted decision tree model. The low-level tracking and calorimeter data are used as inputs in the first method, whereas the second method employs tabular reconstructed object data. The third method uses a Residual Neural Network (ResNet) model based on a CNN architecture. Here, the low-level detector information is converted into three-layer event images with a resolution of 50×50 bins and bin values corresponding to the energy deposits in the electromagnetic and hadronic calorimeters and p_T of tracks observed in the tracking system. We proceed to estimate the expected p-values for each hypothesis, for a given number of observed events in the signal region, originating from each of the possible scenarios.

The rest of the paper is organised as follows. In Sect. 2 our simulation setup and event selection are explained. The six model hypotheses used in this study are also given in this section. The distributions of several observables are studied in Sect. 3. The three ML models used in this study are laid out in Sect. 4. Our main result, the comparison between different ML models and the exclusion p-values of the hypothesis test, is shown and discussed in Sect. 5. Section 6 is devoted to conclusions.

2 Monte Carlo simulation and event selection

To study the classification of high-multiplicity events, we first consider semi-classical BH events. For simplicity and to align with the CMS analysis [48], we focus on non-rotating BHs and fix the fundamental Planck scale at $M_* = 4$ TeV throughout our analysis. Within this assumption, the two most important parameters are the number of extra dimensions, n, and the minimal BH mass, M_{min}, which we vary. The minimal BH mass is defined as the partonic threshold energy, where the semi-classical BH events are turned on. Such a phenomenon is expected under the Hoop conjecture.[3] We treat the production cross section as an independent parameter, rather than relying on specific assumptions to connect it with other parameters. The observed number of events is therefore not used for the classification input. In this treatment, the model-dependent CMS bounds, $M_{\text{min}} \gtrsim 10$ TeV, [48] is also not applicable.

The BH events are generated at parton level using the `BlackMax` event generator [57]. `BlackMax` allows for the choice of the number of extra dimensions, n, and the minimal BH mass, M_{min}. In this study, we examine five different BH hypotheses with number of extra dimensions set to $n = 2$, 4 and 6 and minimal BH masses to $M_{\text{min}} = 8$, 10, and 12 TeV. In the semi-classical regime, the BH loses its mass via Hawking radiation. However, when the BH mass reaches M_*, the semi-classical approximation breaks. The detailed mechanism of the final burst of the BH evaporation is not known. The `BlackMax` takes the minimal approach. Once the BH mass reaches M_*, the BH decays into the minimal number of SM particles, which conserves all unbroken charges, such as the total electric charge, colour and four-momentum. The back reaction from this burst process to the BH geometry is not considered. Altering these assumptions might affect our conclusion quantitatively.

[3] As opposed to this case, there is a scenario in string theory, where semi-classical BH-like events turn on smoothly as a function of the partonic collision energy [56]. We do not consider such a scenario.

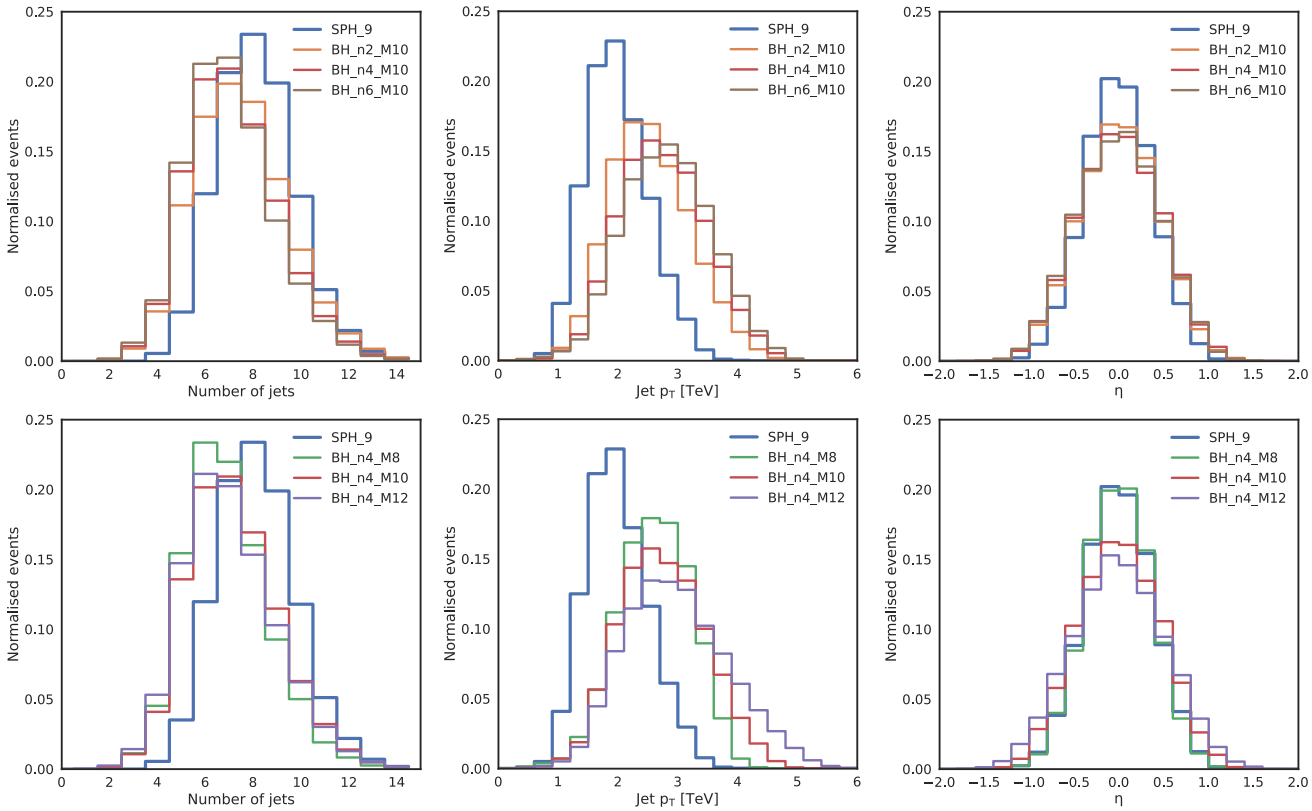

Fig. 1 Normalised distributions of the number of signal jets (left), the p_T of all signal jets (middle) and the pseudorapidity η of the highest-p_T jets (right). The sphaleron events are shown with blue histograms. The plots in the upper panel show the BH scenarios with $n = 2$ (orange), 4 (red) and 6 (brown), while those in the lower panel show the BH scenarios with $M_{\text{min}} = 8$ (green), 10 (red) and 12 (purple) TeV

The EW sphaleron/instanton-induced events are generated using `Herwig 7` [58–64] with the instanton addon library [65]. We fix the threshold energy of the sphaleron production at $E_{\text{thr}}^{\text{sph}} = 9$ TeV. Both sphaleron and black hole events are simulated assuming proton–proton collisions at $\sqrt{s} = 13$ TeV. It has been argued that sphalerons may be dominantly produced in association with multiple EW bosons [26,31]. This argument is, however, based on an approximation in a low energy regime, $E \ll E_{\text{thr}}^{\text{sph}}$. We do not know whether this conclusion still holds at $E \sim E_{\text{thr}}^{\text{sph}}$.[4] We found that the sphaleron events are most similar to the BH events when the boson production is switched off (see the first plot in Fig. 1). If the multi-boson production is included, sphaleron events would have much higher jet multiplicities than BH events and the discrimination between them would be trivial. We therefore choose the zero boson option to study the most interesting case for the discrimination problem. For the parton shower and hadronisation, we employ `Herwig 7` for both BH and sphaleron events. The detector response and object reconstruction are simulated with the fast multi-purpose detector response simulation framework `Delphes 3` [66], assuming the conditions and geometry of the ATLAS detector. This paper is a proof-of-concept study and does not include any real ATLAS data or ATLAS simulations. Instead, we use computationally inexpensive open-source tools that are good approximations of, not openly available, ATLAS internal simulations. Although we do not expect a large effect from the detector modelling, a detailed study using the full detector simulation may be necessary for the real-world application.

The six model hypotheses studied in this paper are listed in Table 1. In both BH and sphaleron scenarios, we treat the production cross sections as unknown parameters, since we wish to focus on discrimination with kinematics only, while keeping in mind, however, that the current experimental limit on the sphaleron and BH cross sections is $\sigma \lesssim 0.1$ fb [42,48] with $E_{\text{thr}}^{\text{sph}}$, $M_{\text{min}} \sim 8$–12 TeV.[5]

[4] In fact, the CMS analysis [48] is based on sphaleron events without the boson production.

[5] One might wonder whether one can treat the signal cross section as an independent parameter from the partonic threshold energy, $M_{\text{thr}} = E_{\text{thr}}^{\text{sph}}$ or M_{min}. Such a treatment is indeed possible because the partonic cross section is phenomenologically modeled as $\hat{\sigma}(\hat{s}) = \hat{\sigma}_0 \Theta(\sqrt{\hat{s}} - M_{\text{thr}})$ [44,48], where $\Theta(x)$ is the Heaviside step function. The hadronic

Table 1 List of the six model hypotheses used in this study. The first row is the sphaleron scenario with the 9 TeV threshold energy, $E_{\rm thr}^{\rm sph}$. The other five rows correspond to the BH scenarios with different number of extra dimensions n and the minimal BH masses $M_{\rm min}$. The signal efficiencies $\epsilon_{\rm sig}$ for the signal region ($N \geq 5$, $S_T \geq 7$ TeV) are shown in the last column

Hypothesis	n	$M_{\rm min}$	$E_{\rm thr}^{\rm sph}$	$\epsilon_{\rm sig}$
SPH_9	–	–	9 TeV	0.18
BH_n4_M8	4	8 TeV	–	0.17
BH_n2_M10	2	10 TeV	–	0.55
BH_n4_M10	4	10 TeV	–	0.49
BH_n6_M10	6	10 TeV	–	0.43
BH_n4_M12	4	12 TeV	–	0.65

Before delving into the classification problem, we would like to highlight that the SM background is under control. In particular, in Ref. [48], the CMS collaboration has performed an analysis to search for semi-classical BHs and EW sphalerons and has demonstrated that the SM background, dominantly QCD multijet events, can be heavily suppressed with an appropriate event selection. In the CMS analysis, the event selection is based on the two reconstructed quantities, N and S_T. The variable N is defined as the number of high-p_T objects in an event, with $p_T > 70$ GeV, including jets with $|\eta| < 5$, isolated electrons and photons with $|\eta| < 2.5$, and isolated muons with $|\eta| < 2.4$. In the following, we call the reconstructed objects satisfying the above criteria, *signal objects*. For example, signal electrons comprise of the isolated electrons with $p_T > 70$ GeV and $|\eta| < 2.5$. The S_T is the scalar sum of the missing transverse energy $p_T^{\rm miss}$ and the magnitude of transverse momenta of all N signal objects, $S_T \equiv |p_T^{\rm miss}| + \sum_{i=1}^{N} |p_T^i|$. The CMS study has shown that the SM background is reduced to $\sim$ 0.1 event at 36 fb^{-1} (13 TeV) by demanding $S_T \geq 7$ TeV and $N \geq 5$. On the other hand, 17–65% of the signal events remain after this event selection, as shown in Table 1. In the following analysis, we apply essentially the same event selection. The only difference is that our signal objects are defined with a tighter pseudorapidity cut, $|\eta| < 2.4$.[6] Since the SM background can be safely neglected by this selection, we do not consider it further in our analysis. Assuming the signal cross section is just below the current limit and the event selection efficiency is $\sim$ 50%, we expect around 30 signal events after the cut at the LHC run-3 (13.6 TeV, 300 fb^{-1}), while the number of expected backgrounds is $\sim$ 1 or smaller since our selection cuts are tighter than the CMS ones.[7] At the High-Luminosity LHC (HL-LHC), with an integrated luminosity of 3 ab, one has to even tighten the selection cuts to suppress the SM background below 1. This is necessary to justify our analysis where the SM background is not included in the simulation. After such a tighter selection cut, one can still expect the observed signal events of $\mathcal{O}(100)$ or more, especially if the HL-LHC is operated with the collision energy of 14 TeV. Although we envisage the collision energy slightly higher than 13 TeV, our simulation is performed with 13 TeV throughout, for simplicity. We, however, do not expect significant changes in the event kinematics and our qualitative result is valid for collision energies between 13 and 14 TeV. Since we do not expect a large number of observed signal events, our goal is to find a model discrimination method applicable to low signal statistics, $\mathcal{O}(30) - \mathcal{O}(100)$.

3 Kinematical distributions

In order to assess the sensitivity to the model hypotheses of high-level kinematical variables, we compare their distributions in Figs. 1 and 2. In the upper panels of both figures we fix the minimal BH mass at 10 TeV and vary the number of extra dimensions as $n = 2$ (orange), 4 (red) and 6 (brown), whereas in the lower panels, we fix $n = 4$ and varied $M_{\rm min} = 8$ (green), 10 (red) and 12 (purple) TeV. In both panels, the distributions for sphaleron events are shown in blue.

The left panel of Fig. 1 shows the number of signal jets, where we see that the sphaleron has on average 2–3 additional signal jets compared to the BH scenarios. The majority ($\sim$ 90%) of sphaleron events have 6–10 signal jets and the distribution peaks at 8. On the other hand, the distributions for BH scenarios peak around 5–6 jets and have a relatively long tail towards higher multiplicities. The signal jet multiplicity is almost insensitive to the number of extra dimensions and the minimal BH mass.

The middle panel of Fig. 1 displays the accumulated p_T distributions of all signal jets. We see that the average signal jet p_T is much lower for sphaleron events. The distribution for sphaleron events has a sharper peak around 2 TeV, whereas the distributions for BH scenarios have a broader shape, and the average jet p_T is much higher, around 2.5–3.5 TeV. We observe in the upper middle plot that the jet p_T for $n = 2$ BH is slightly lower than that for $n = 4$ and 6, on average. In the lower middle plot, we also see that the jet p_T distributions

Footnote 5 continued
cross section, obtained by integrating $\hat{\sigma}(\hat{s})$ with the parton distribution functions, still depends on unknown parameter σ_0, even after fixing $M_{\rm thr}$.

[6] We have checked the change of the signal efficiency by this modification is less than 1%.

[7] Here we took into account the fact that by the energy increase from 13 to 13.6 TeV, the signal cross section is roughly doubled for $E_{\rm thr}^{\rm sph}$, $M_{\rm min} \sim 9$ TeV [37], while the background cross section does not change significantly.

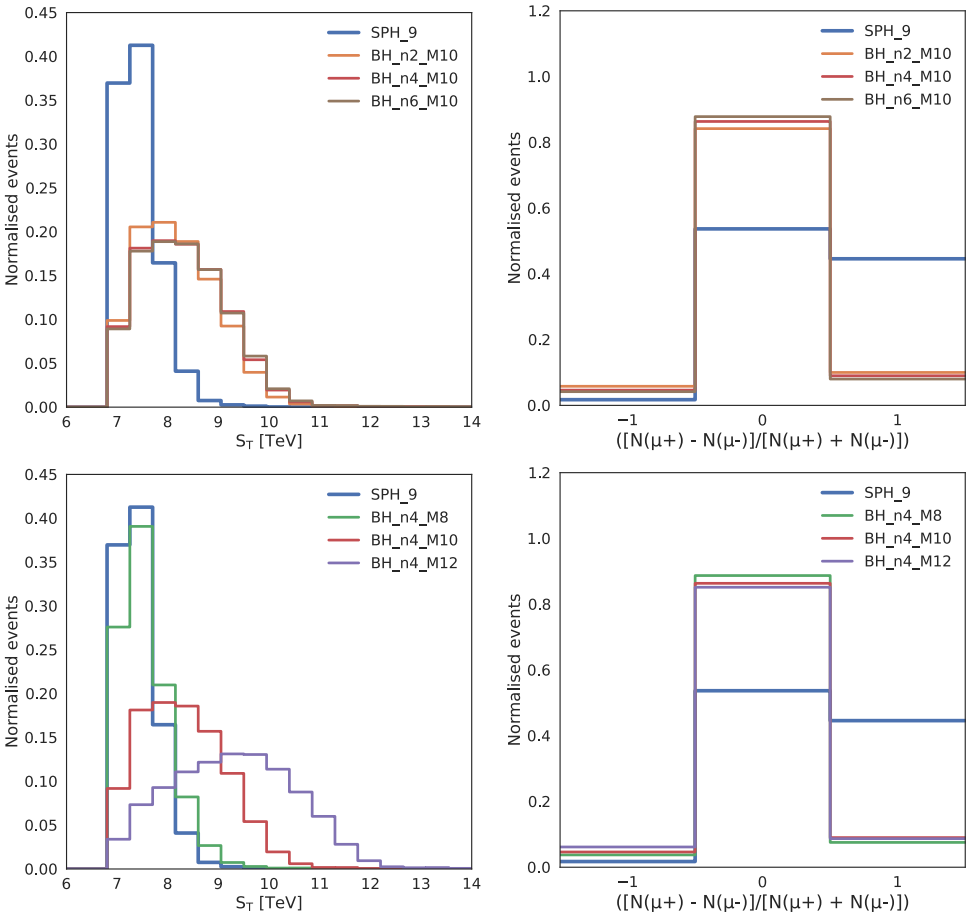

Fig. 2 Normalised distributions of S_T (left) and the muon charge asymmetry CA_μ (right). The same colour coding is used as in Fig. 1

have longer tails in the higher-p_T region for larger minimal BH masses.

The distributions of the pseudorapidity η of the highest-p_T jet are shown in the right panel of Fig. 1. As can be seen, the most energetic jet is produced in the central region, $|\eta| \lesssim 1$, both for sphaleron and BH events. The η distributions are similar among all scenarios, though one can observe that the sphaleron and the $(M_{\min}, n) = (8\,\mathrm{TeV}, 4)$ BH have slightly narrower distributions, compared to the other BH scenarios.

The left panel of Fig. 2 shows the S_T distributions. As can be seen, the S_T distribution for sphalerons sharply peaks around 7.5 TeV. We observe that the S_T distribution for BH scenarios is sensitive to $M_{\min}$, but not to the number of extra dimensions, n. We see in the lower-left plot that the S_T distribution for the $(M_{\min}, n) = (8\,\mathrm{TeV}, 4)$ BH is similar to that for sphalerons. The distribution gets broader and shifts towards higher energies as the minimal BH mass increases. The S_T distribution peaks around 9.5 TeV for the $(M_{\min}, n) = (12\,\mathrm{TeV}, 4)$ BH scenario.

The right panel of Fig. 2 displays the charge asymmetry of signal muons, defined by the difference between the numbers of μ^+ and μ^- in an event, divided by the total number of muons: $\mathrm{CA}_\mu \equiv (N_{\mu^+} - N_{\mu^-})/(N_{\mu^+} + N_{\mu^-})$. If the event contains no signal muons, we set $\mathrm{CA}_\mu = 0$. We see in the plots that there is a large excess of μ^+ in sphaleron events: sphaleron events have $\mathrm{CA}_\mu = 0$ or 1 with $\sim 50\%$ chance. This asymmetry is expected because in the quark-quark initial state, the sphaleron vertex involves exactly one anti-muon or exactly one anti-muon-neutrino. Extra (anti-)muons may arise from decays of anti-top-quark or heavy hadrons. Contrary to the sphaleron case, the majority ($\sim 90\%$) of BH events have zero muon charge asymmetry. This is expected because semi-classical BH decays are essentially a thermal process. In the plots, we however observe a small excess for μ^+ in all BH samples. This is a consequence of the charge conservation of BH decays and the proton contains more up-quarks than down-quarks, especially in a large-x region of the parton distribution function. Although the muon charge asymmetry has different distributions between BH and EW sphaleron events, it is not the best discrimination variable, particularly when the available data size is small, as the majority of events have $\mathrm{CA}_\mu = 0$.

Observed differences in several kinematical distributions between sphaleron and various BH scenarios suggest that the separation between sphaleron and BH scenarios should be relatively straightforward with sufficient collider data, though the optimal discrimination method is not clear. Separation between different $M_{\min}$ among BH scenarios seems

also possible with some resolution, since the jet p_T and S_T distributions depend, to some extent, on M_{min}. On the other hand, separation of the number of extra dimensions n seems challenging as no distribution in Figs. 1 and 2 exhibits a clear dependence on n. In the following sections, we investigate how well one can distinguish different scenarios, with a given number of observed events, using modern ML techniques.

4 Machine learning setup

In this study, we examine three types of ML methods. The first two are based on the well-established XGBoost algorithm [67], which creates an ensemble of decision trees to effectively separate different events. The difference between the two models based on XGBoost is the choice of input features. In the first approach, we use high-level variables based on reconstructed objects, while in the second we utilise low-level tracker and calorimeter signals. The third method uses the CNN-based ResNet [68] architecture, which is known to be very powerful in image recognition applications.

The ML models are all trained and evaluated using shared input data consisting of 10000 (training), 3000 (validation) and 15000 (testing) events for each physics model hypothesis, giving a total of 60000, 18000 and 90000 events, respectively. These are the number of events after cuts are applied, so the models are trained on a balanced dataset. The models are trained using the training set, fine-tuned using the validation set and finally, the networks are used to predict the labels of the independent test set events. The predicted and true labels of the test set are then used to calculate the metrics for evaluation. For this study, the metric chosen is a simple global accuracy (ACC), i.e., the number of correctly labelled events divided by the total number of events, such that perfect labelling of all events corresponds to ACC $= 1.0$, and a random classifier results in ACC $= 1/n$, where n is the number of distinct classes (under the assumption that the dataset is balanced, i.e., all classes are equally represented).[8]

To assess the uncertainties due to the randomness of the ML training, for each of the three ML methods, we create five independent classifiers (decision trees for XGBoost and neural networks for ResNet) by shuffling the training data at the start of each training. The mean global accuracy and its standard deviation are estimated with these five independent classifiers.

4.1 XGBoost

The XGBoost algorithm was developed during the highly successful Higgs ML 2014 Kaggle competition (http://www.kaggle.com/c/higgs-boson) [70]. It is a regularizing gradient boosting framework library and enables fast training of decision tree models. XGBoost is implemented in many modern HEP analyses such as the diphoton search from ATLAS [71]. As such, XGBoost is a state-of-the-art machine-learning tool for high-energy particle physics and is expected to yield good results in a short training time.

In this study, XGBoost models with two different input types have been explored. The first XGBoost method (XGBoost-High) uses the high-level inputs of the reconstructed objects. The selected input features are[9] p_T. the first eight jets (p_T^{miss}, η, ϕ), the first two electrons and muons (p_T, η, ϕ) and the missing transverse energy (p_T^{miss}, ϕ), which comprises 38 input features. The XGBoost algorithm tolerates missing values, but the data must be the same length for all events. The data are padded with `NaN` if an event does not have enough jets or leptons.

In the second XGBoost method (XGBoost-Low) low-level input is used. In this case, the input features consist of the signals in the tracking system, the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). The calorimeter hits, also referred to as calorimeter towers, represent the summed-up energies at discrete angular positions across the detector. This means that the radial coordinates where the energy was deposited are not retained. The 30 highest-energy hits in each of the three different detector systems, and their corresponding angular position, were selected. This sums to 270 input features in total. Although we experimented with other variations of input features, we did not find any improvement by adding more features with lower-energy hits.

4.2 Convolutional neural network

CNNs are neural networks with convolutional layers. It is known that this architecture works extremely well in recognizing patterns in image-type data, that is data which have a matrix structure where the relative positioning of each input in the matrix is important. In this study, we use the ResNet18 architecture with 18 layers,[10] implemented with the PyTorch package [72]. We convert low-level collider event data into event images as follows: starting from the 2D angular coordinates of the calorimeter towers and tracks, (ϕ, η), we divide the range $\phi \in [-\pi, \pi]$, $\eta \in [-5, 5]$ into 50×50 equal sized bins.[11] The total energy deposit (in GeV, not normalised) in the ECAL, HCAL and tracks in a given bin is assigned to their respective layer in that bin. The result is a 3-layer 2D

[8] Code, results and data specifications are openly available on our project GitHub https://github.com/choisant/imcalML [69].

[9] The jets and leptons are ordered by descending.

[10] We tested models with more layers, but did not find an increase in performance.

[11] Higher resolutions than 50×50 were tested. We found either no improvement or slightly worse performance.

Table 2 The global accuracy (in percent) as well as the training and testing time for XGBoost-High, XGBoost-Low and ResNet18 ML methods. The mean and error of the global accuracy were estimated from an ensemble of five separately trained classifiers for each ML method. The quoted training and testing times were measured based on a GPU with an NVIDIA RTX A4500 graphic card

Model	XGBoost-High	XGBoost-Low	ResNet18
Accuracy	48.91 ± 0.05	46.69 ± 0.08	$\mathbf{57.92 \pm 0.37}$
Training time	**2 min**	3.5 min	8.5 min
Testing time	10 s	**7 s**	31 s

The values highlighted with bold indicate the best performance among the three architectures

histogram with the input shape $50 \times 50 \times 3$, which we refer to as an event image and has 7500 input features in total. In the appendix, we show example event images of each of the physics scenarios in Fig. 5.

In the course of the training process, data augmentation is applied to the images in order to artificially increase the statistics of the dataset. The augmentation consists of random rotations in the ϕ-direction and random flipping of the η axis. These transformations do not affect the physics of the events, because of the symmetries of the detector and the collision setup. The images in the validation and testing set are not transformed in any way.

During the training, random batches of 128 images are loaded into memory at a time. Data augmentation is applied to each image, and then the network is trained using an exponential cyclic learning rate scheduler [73] with a base learning rate of 0.001, a maximum learning rate of 0.01, and gamma = 0.85 and the Adam optimization algorithm [74] based on the calculated cross-entropy loss from the batch. The network is trained for 40 epochs.

5 The classification performance

The mean global accuracy with uncertainties due to the stochastic ML training process, as well as training and testing times for each ML method, is presented in Table 2. The three ML methods we examined have global accuracy classifier scores ranging from 46.69 ± 0.08 (XGBoost-Low) to 57.92 ± 0.37 (ResNet18). It is interesting that ResNet18, which uses only low-level data, can perform significantly better than XGBoost does with high-level data inputs. It should be noted that the training and testing times for the CNN model are approximately 3–4 times longer than for XGBoost. This is a natural effect of the deep neural network's complex structure and the large difference in the number of input features.

Figure 3 shows the confusion matrices of the three ML methods we examined. The separation of sphalerons and BHs is the objective that all models handle well. In particular, ResNet can correctly identify the sphaleron events with $\sim 91\%$ accuracy. Furthermore, when separating the different BH samples, the models are much better at distinguishing samples with different minimal BH masses, than with different numbers of extra dimensions. ResNet correctly identifies 92% of all the black holes with a minimum mass of 12 TeV, compared to 71% and 80% respectively for XGBoost-High and XGBoost-Low. ResNet is also fairly good at separating the different number of dimensions. It correctly classifies 49% of the events with 2 extra dimensions and 42% of the ones with 6 extra dimensions.

To demonstrate the practical application of the ML classifiers, we perform hypothesis tests using Poisson statistics, for a given number of observed signal events, N_{obs}, in our signal region ($N \geq 5$, $S_T \geq 7$ TeV). We have conducted this exercise only with the ResNet model as it outperforms XGBoost.

For each observed event i of model hypothesis scenario J,[12] ResNet assigns a classification label, L_i, where J and L_i belong to the set of six model hypotheses in Table 1. Using the test set of 15000 events, we first create a normalised template distribution of classification labels L for each scenario J. Roughly speaking, each bin of the distribution, $P_J(L)$, can be interpreted as the probability that an observed event of scenario J is classified as scenario L by the ResNet.

Iterating over all the possible scenarios where the data corresponds to hypothesis I, we test each hypothesis J to see if it can be excluded. We define the label L_J^* into which the event of J is most likely to be classified by the ResNet: i.e., $P_J(L_J^*) \geq P_J(L)$ for all L. If N_{obs} events are observed in the signal region, the average number of events labelled as L_J^* is $\lambda = N_{\text{obs}} \cdot P_J(L_J^*)$. If hypothesis J is correct, the probability of observing n events labelled as L_J^* is given by the Poisson function

$$P(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}. \tag{5.1}$$

The hypothesis J will be excluded at 95% confidence level if the number of observed events labelled as L_J^*, $N(L_J^*)$, in the data is too few: i.e.,

$$p_J \equiv \sum_{n=0}^{N(L_J^*)} P(n|\lambda) < 0.05. \tag{5.2}$$

Given the current limits on the BH and EW sphaleron cross sections, [42,48], $\sigma \lesssim 0.1$ fb, we do not expect that a large number of BH/sphalerons events could be collected in LHC Run-2, -3 and HL-LHC. For a small number of signal events, the corresponding p-values can fluctuate significantly. To understand the magnitude of the fluctuation, we performed 3000 pseudo-experiments, sampled from the 15000 event test

[12] There is no case in which the real LHC data could contain a mix of scenarios.

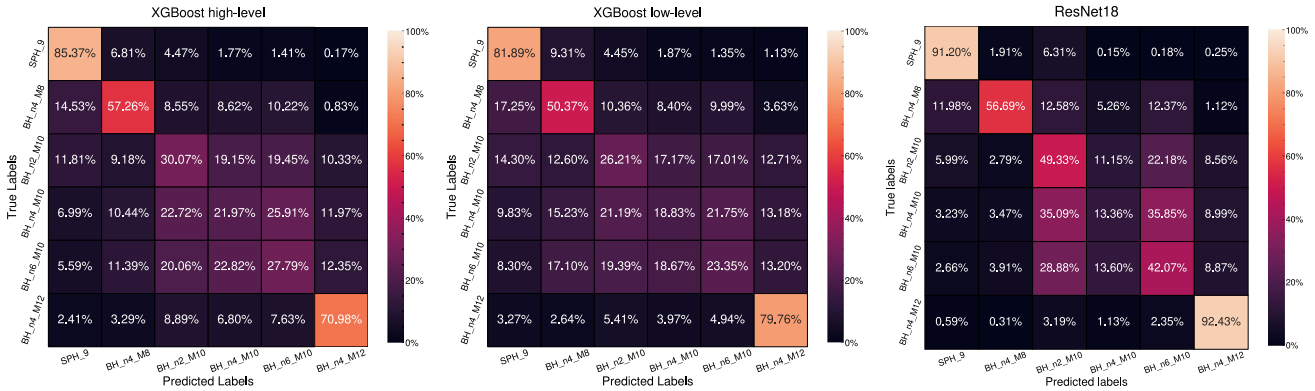

Fig. 3 Confusion matrices for the XGBoost-High (left), XGBoost-Low (middle) and CNN ResNet18 (right)

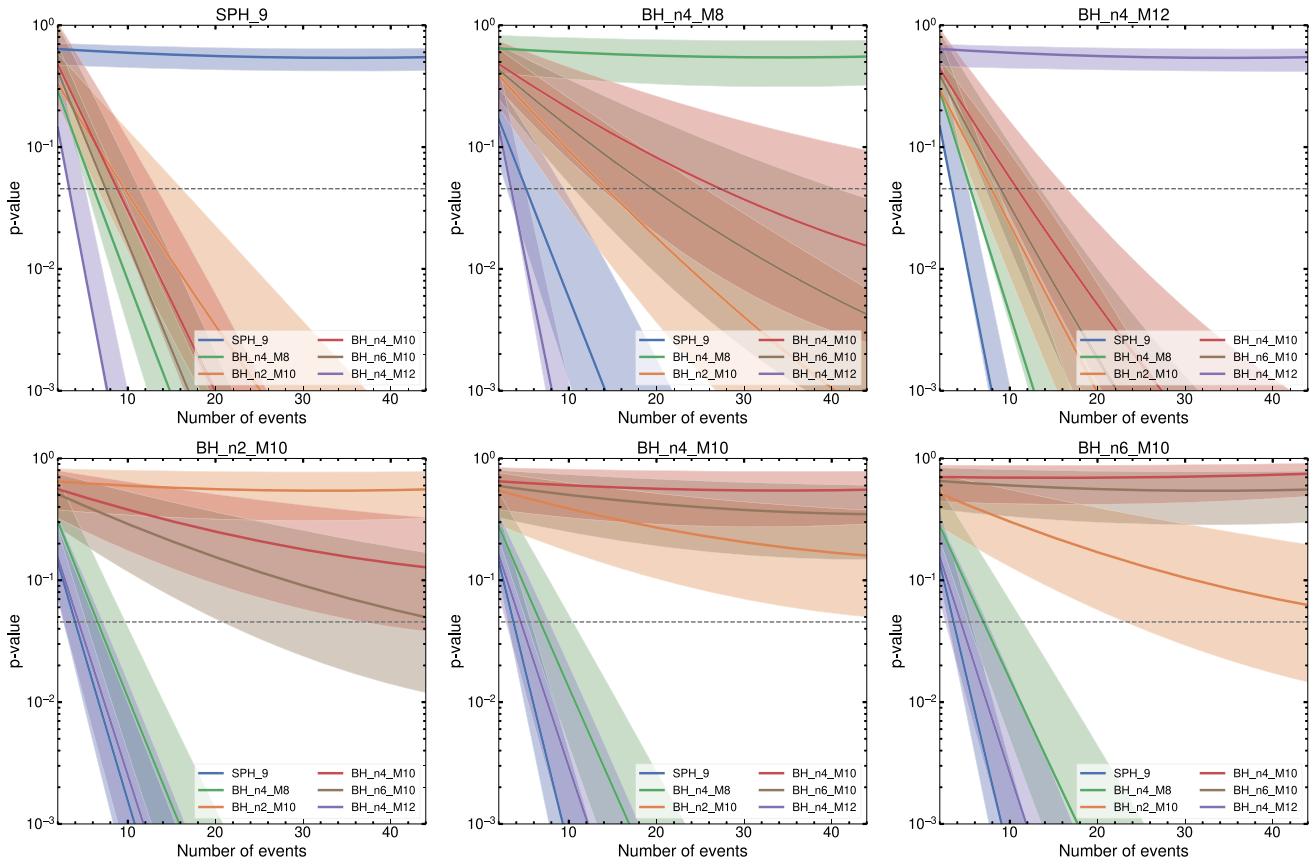

Fig. 4 The exclusion p-values of six model hypotheses calculated using Poisson statistics: SPH_9 (blue), BH_n4_M8 (green), BH_n2_M10 (orange), BH_n4_M10 (red), BH_n6_M10 (brown), BH_n4_M12 (purple) as a function of the signal events registered in the signal region. The true hypotheses are indicated at the top of each plot. The horizontal black-dashed line represents the 2-σ exclusion

set, for each N_{obs} of the "true" scenario and estimated the upper and lower errors on the p-values.

In Fig. 4, we show the result of our p-value estimation as a function of the number of observed events in the signal region. In the top-left plot, the true scenario (the origin of signal events) is SPH_9 (EW sphaleron scenario with 9 TeV threshold energy). The blue curve and the band around it show the p-value for the SPH_9 hypothesis and the standard deviation, estimated from the pseudo-experiments, respectively. As expected, the p-value is $\mathcal{O}(1)$ and does not decrease as the number of signal events increases, since it is the correct hypothesis. The other curves show the p-values for various BH hypotheses: BH_n4_M8 (green), BH_n2_M10 (orange), BH_n4_M10 (red), BH_n6_M10

(brown), BH_n4_M12 (purple). The horizontal dashed black line in the plot corresponds to the 2-σ ($p \simeq 0.05$) exclusion. Observing less than 15 events is often enough to exclude all of our BH hypotheses at more than the 2-σ level.

The top-middle and top-right plots show the p-values of various hypotheses when the true scenarios are BH_n4_M8 and BH_n4_M12, respectively. In both cases, we can observe that the p-values for the correct hypotheses remain $\mathcal{O}(1)$ for a larger number of signal events. We also notice that the sphaleron hypothesis and the BH hypothesis with M_{min} differ by 4 TeV from the correct hypothesis (i.e., BH_n4_M12 (purple) in the middle plot and BH_n4_M8 (green) in the right plot), are excluded very quickly while accumulating signal events. For those scenarios, a small number of events are sufficient for a 2-σ exclusion. In the same plots, we see the p-values for the BH hypotheses whose M_{min} differs by 2 TeV from the correct hypothesis. These are the scenarios with $M_{\text{min}} = 10$ TeV and $n = 2$ (orange), 4 (red) and 6 (brown). We see that the p-values of these three hypotheses decrease at a somewhat slower rate.

The three plots at the bottom of Fig. 4 show the p-values when the true scenarios have $M_{\text{min}} = 10\,\text{TeV}$ and $n = 2$ (bottom-left), 4 (bottom-middle) and 6 (bottom-right). The sphaleron hypothesis can be excluded with a small number of events at the 2-σ levels. The exclusion of the BH scenario with $M_{\text{min}} = 12$ TeV (purple) is easier than that with $M_{\text{min}} = 8\,\text{TeV}$ (green). As can be seen from these three plots, the discrimination between the same M_{min} and different n is more challenging. 2-σ discrimination between $n = 2$ and $n = 6$ requires ~ 45 events.

6 Conclusions

Large-multiplicity final states of jets and leptons at the LHC are ubiquitous in exotic processes, both within the Standard Model and within its extensions. In particular, final states with $\mathcal{O}(10)$ jets plus a few leptons are expected in thermal decays of semi-classical Black Holes (BHs), which are anticipated to be produced in models with large extra dimensions, characterised by two main parameters: the number of extra dimensions and the minimal mass of the BH. Similar final states are also expected in the EW sphaleron/instanton-induced processes. In this study, we have investigated whether one can discriminate among the sphaleron scenario and five different BH scenarios listed in Table 1, through events collected at the LHC, using modern Machine Learning (ML) methods.

In studying kinematical distributions we observed that sphalerons have significantly different distributions than those of various BH scenarios, in the number of jets, p_T of jets, as well as the muon charge asymmetry. The S_T and jet p_T distributions also exhibit a dependence on the minimal BH masses. Despite these differences, the optimal model separation method, which works for low signal statistics, is unclear due to the complexity of large-multiplicity final states. On the other hand, no clear sensitivity to the number of extra dimensions was found in any of the distributions shown in Figs. 1 and 2.

In this study, three ML models have been examined: XGBoost with low- and high-level inputs (XGBoost-Low and XGBoost-High) and a convolutional neural network ResNet model. In the latter, the input was the low-level detector information, converted into three-layer binned event images, corresponding to the signals in the ECAL, HCAL and tracking system. We found that the discrimination power is the highest for the ResNet model as it outperforms the state-of-the-art XGBoost method by achieving a 9% higher global accuracy score on our test set. This is quite remarkable and highlights the capacity of neural networks to utilize low-level, high-dimensional data.

To assess the practical applicability of the best-performing ML classification, we evaluated the exclusion p-value of each hypothesis, J, for a given number of observed signal events (in the signal region), originating from the true scenario I. We demonstrated that the sphaleron hypothesis can be discriminated from various BH scenarios with a small number of events. Separation between BH scenarios with different minimal BH masses is also possible with a reasonable number of events collected at LHC Run-2, -3 and HL-LHC. The discrimination among the BH scenarios with the same minimal BH mass but a different number of extra dimensions is more challenging and requires a larger number of signal events, which may be collected at future high-energy colliders.

Finally, we emphasise that the work presented in this paper should be considered a proof-of-concept study, as it entirely relies on MC simulations. In the actual experimental situation, one must validate and correct the MC modelling with the real data before using it for model discrimination. Such a process is non-trivial and interesting in its own right. We, however, postpone this subject for future work.

Acknowledgements The research of A.S.G, F.K., A.L., R.M., K.S. and T.S leading to these results has received funding from the Norwegian Financial Mechanism for years 2014-2021, grant no DEC-2019/34/H/ST2/00707. T.S. acknowledges support from the Research Council of Norway, grant no 314472, which is supporting the research of A.S.G and I.S. K.S. is supported by the National Science Centre, Poland, under research grant 2017/26/E/ST2/00135. R.M. is supported by the National Science Centre, Poland, under PRELUDIUM research grant no 2021/41/N/ST2/00972 and under SONATA BIS grant no 2020/38/E/ST2/00243. A.P. acknowledges support by the National Science Foundation under Grant No. PHY 2210161.

Funding This study was funded by Norwegian Financial Mechanism (DEC-2019/34/H/ST2/00707), National Science Foundation (PHY 2210161), Norges Forskningsråd (314472) and Narodowe Centrum Nauki (2017/26/E/ST2/00135, 2021/41/N/ST2/00972, 2020/38/E/ST2/00243).

Data Availability Statement This manuscript has associated data in a data repository. [Authors' comment: This manuscript has associated open data and metadata available both in Zenodo (https://zenodo.org/doi/10.5281/zenodo.10033265) and GitHub (https://github.com/choisant/imcalML).].

Code availability My manuscript has associated code/software. [Author's comment: The code used in this study is publicly available both in Zenodo (https://zenodo.org/doi/10.5281/zenodo.10033265) and GitHub (https://github.com/choisant/imcalML).].

Funded by SCOAP[3].

Appendix A: Example event images

We show below in Fig. 5, event images of various physics scenarios: SPH_9, BH_n4_M8, BH_n2_M10, BH_n4_M10, BH_n6_M10 and BH_n4_M12. Three randomly-chosen events in the validation dataset are used for each scenario as examples. The signals in the ECAL, HCAL and tracking system correspond to the red, green and blue colour intensities, respectively. To aid visualisation, we map the accumulated energy deposits (p_T of the tracks), E_i, in bin i to the colour intensity $I_i = f(E_i)$ with a conversion function $f(E_i) = \arctan(\ln(E_i/20\,\text{GeV}))/\pi + \frac{1}{2} \in [0, 1]$. The reconstructed jet (anti-k_T jet with radius parameter $R = 0.4$) positions for each event are indicated with orange circles of radius 0.4.

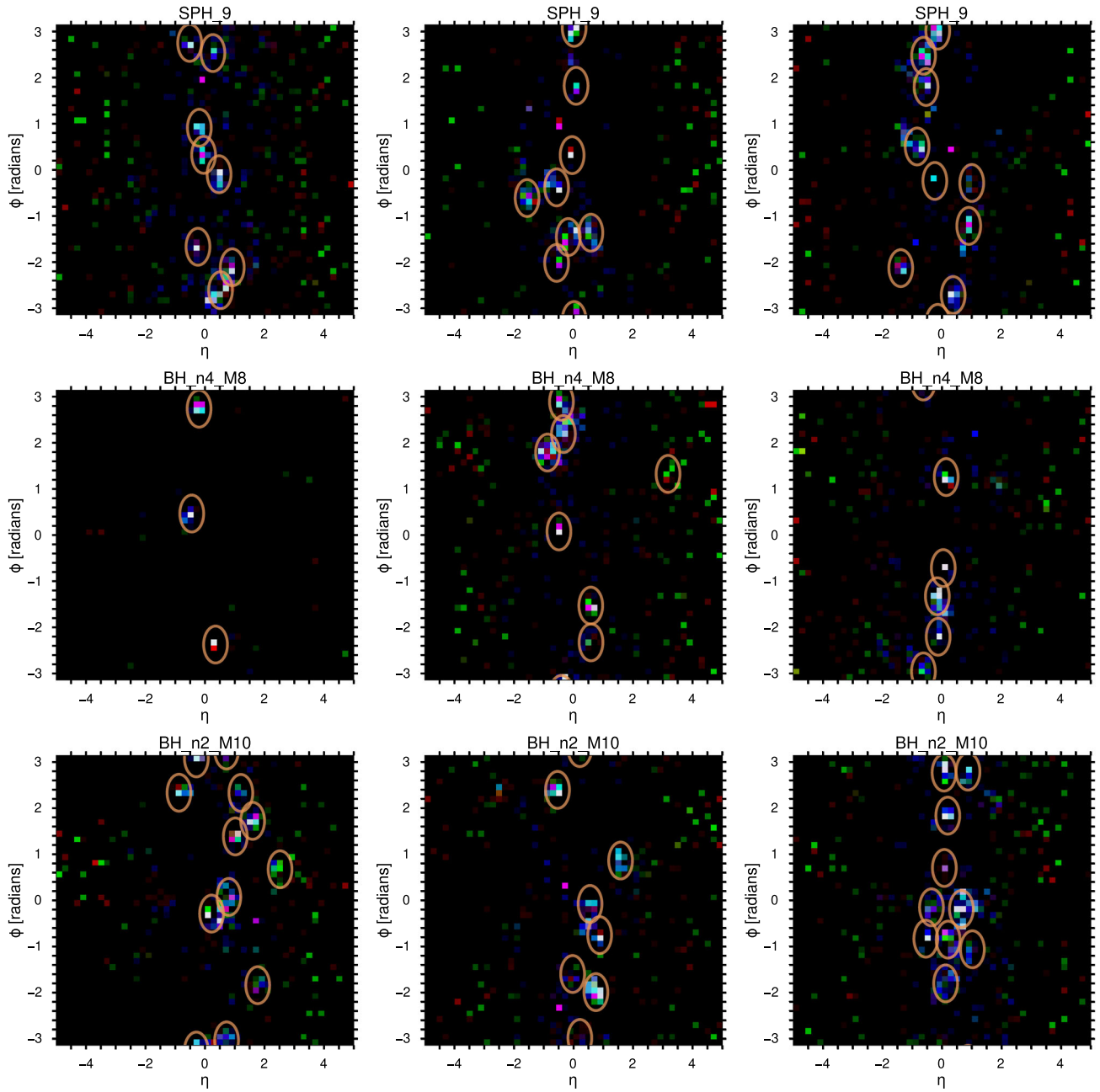

Fig. 5 Event images of various physics scenarios. Three randomly-chosen events are shown for each scenario. The red, green and blue colour intensities correspond to the signals in the ECAL, HCAL and tracking system, respectively. The reconstructed jet positions for each event are indicated with orange circles of radius 0.4

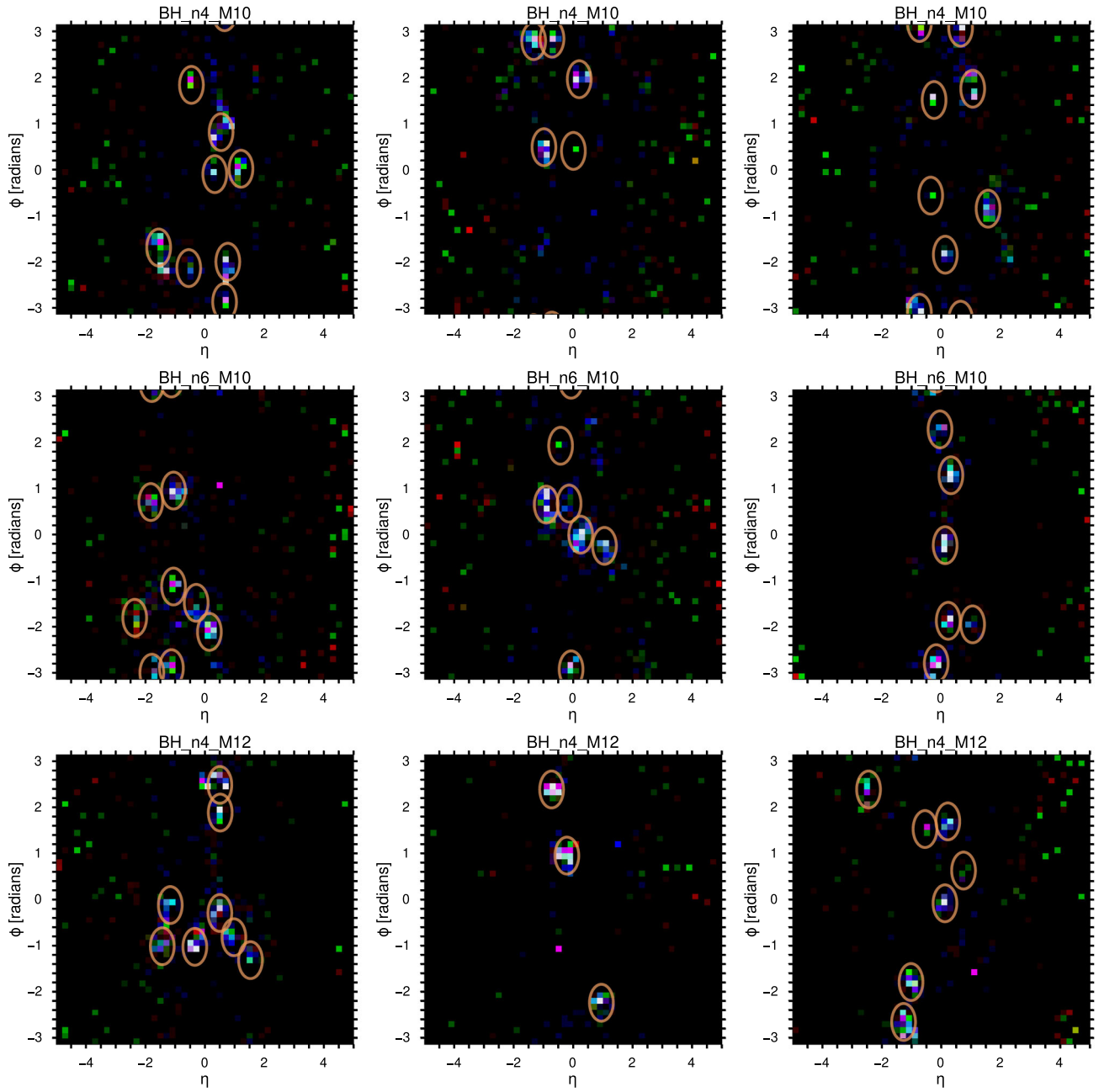

Fig. 5 continued

References

1. N. Arkani-Hamed, S. Dimopoulos, G.R. Dvali, The hierarchy problem and new dimensions at a millimeter. Phys. Lett. B **429**, 263–272 (1998). https://doi.org/10.1016/S0370-2693(98)00466-3. arXiv:hep-ph/9803315
2. I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos, G.R. Dvali, New dimensions at a millimeter to a Fermi and superstrings at a TeV. Phys. Lett. B **436**, 257–263 (1998). https://doi.org/10.1016/S0370-2693(98)00860-0. arXiv:hep-ph/9804398
3. N. Arkani-Hamed, S. Dimopoulos, G.R. Dvali, Phenomenology, astrophysics and cosmology of theories with submillimeter dimensions and TeV scale quantum gravity. Phys. Rev. D **59**, 086004 (1999). https://doi.org/10.1103/PhysRevD.59.086004. arXiv:hep-ph/9807344
4. T. Banks, W. Fischler, A model for high-energy scattering in quantum gravity. arXiv:hep-th/9906038
5. S. Dimopoulos, G.L. Landsberg, Black holes at the LHC. Phys. Rev. Lett. **87**, 161602 (2001). https://doi.org/10.1103/PhysRevLett.87.161602. arXiv:hep-ph/0106295
6. D.M. Eardley, S.B. Giddings, Classical black hole production in high-energy collisions. Phys. Rev. D **66**, 044011 (2002). https://doi.org/10.1103/PhysRevD.66.044011. arXiv:gr-qc/0201034
7. H. Yoshino, Y. Nambu, High-energy head-on collisions of particles and hoop conjecture. Phys. Rev. D **66**, 065004 (2002). https://doi.org/10.1103/PhysRevD.66.065004. arXiv:gr-qc/0204060
8. H. Yoshino, Y. Nambu, Black hole formation in the grazing collision of high-energy particles. Phys. Rev. D **67**, 024009 (2003). https://doi.org/10.1103/PhysRevD.67.024009. arXiv:gr-qc/0209003
9. D. Ida, K.-Y. Oda, S.C. Park, Rotating black holes at future colliders: greybody factors for brane fields. Phys. Rev. D **67**, 064025 (2003). https://doi.org/10.1103/PhysRevD.67.064025. arXiv:hep-th/0212108. [Erratum: Phys. Rev. D **69**, 049901 (2004)]
10. C.M. Harris, P. Richardson, B.R. Webber, CHARYBDIS: a black hole event generator. JHEP **08**, 033 (2003). https://doi.org/10.1088/1126-6708/2003/08/033. arXiv:hep-ph/0307305
11. C.M. Harris, M.J. Palmer, M.A. Parker, P. Richardson et al., Exploring higher dimensional black holes at the large hadron collider. JHEP **05**, 053 (2005). https://doi.org/10.1088/1126-6708/2005/05/053. arXiv:hep-ph/0411022
12. G.W. Gibbons, S.W. Hawking (eds.), Particle creation by black holes. Commun. Math. Phys. **43** (1975) 199–220. https://doi.org/10.1007/BF02345020. [Erratum: Commun. Math. Phys. **46**, 206 (1976)]
13. S.B. Giddings, S.D. Thomas, High-energy colliders as black hole factories: the end of short distance physics. Phys. Rev. D **65**, 056010 (2002). https://doi.org/10.1103/PhysRevD.65.056010. arXiv:hep-ph/0106219
14. P. Meade, L. Randall, Black holes and quantum gravity at the LHC. JHEP **05**, 003 (2008). https://doi.org/10.1088/1126-6708/2008/05/003. arXiv:0708.3017
15. CMS Collaboration, Search for resonances and quantum black holes using dijet mass spectra in proton–proton collisions at $\sqrt{s} =$ 8 TeV. Phys. Rev. D **91**, 052009 (2015). https://doi.org/10.1103/PhysRevD.91.052009. arXiv:1501.04198
16. ATLAS Collaboration, Search for quantum black hole production in lepton+jet final states using proton–proton collisions at $\sqrt{s} =$ 13 TeV with the ATLAS detector. Phys. Rev. D. **109**, 032010 (2024) https://doi.org/10.1103/PhysRevD.109.032010
17. CMS Collaboration, Search for heavy resonances and quantum black holes in $e\mu$, $e\tau$, and $\mu\tau$ final states in proton–proton collisions at $\sqrt{s} = 13$ TeV. JHEP **05**, 227 (2023). https://doi.org/10.1007/JHEP05(2023)227. arXiv:2205.06709
18. R.C. Myers, M.J. Perry, Black holes in higher dimensional space-times. Ann. Phys. **172**, 304 (1986). https://doi.org/10.1016/0003-4916(86)90186-7
19. B.S. DeWitt, Quantum field theory in curved space-time. Phys. Rep. **19**, 295–357 (1975). https://doi.org/10.1016/0370-1573(75)90051-4
20. D.N. Page, Particle emission rates from a black hole: massless particles from an uncharged, nonrotating hole. Phys. Rev. D **13**, 198–206 (1976). https://doi.org/10.1103/PhysRevD.13.198
21. N.S. Manton, Topology in the Weinberg–Salam theory. Phys. Rev. D **28**, 2019 (1983). https://doi.org/10.1103/PhysRevD.28.2019
22. F.R. Klinkhamer, N.S. Manton, A saddle point solution in the Weinberg–Salam theory. Phys. Rev. D **30**, 2212 (1984). https://doi.org/10.1103/PhysRevD.30.2212
23. J.C. Taylor (ed.), Pseudoparticle solutions of the Yang–Mills equations. Phys. Lett. B **59**, 85–87 (1975). https://doi.org/10.1016/0370-2693(75)90163-X
24. M.A. Shifman (ed.), On constrained instantons. Nucl. Phys. B **191**, 429 (1981). https://doi.org/10.1016/0550-3213(81)90307-2
25. V.A. Rubakov, M.E. Shaposhnikov, Electroweak baryon number nonconservation in the early universe and in high-energy collisions. Usp. Fiz. Nauk **166**, 493–537 (1996). https://doi.org/10.1070/PU1996v039n05ABEH000145. arXiv:hep-ph/9603208
26. A. Ringwald, High-energy breakdown of perturbation theory in the electroweak instanton sector. Nucl. Phys. B **330**, 1–18 (1990). https://doi.org/10.1016/0550-3213(90)90300-3
27. O. Espinosa, High-energy behavior of baryon and lepton number violating scattering amplitudes and breakdown of unitarity in the standard model. Nucl. Phys. B **343**, 310–340 (1990). https://doi.org/10.1016/0550-3213(90)90473-Q
28. P.G. Tinyakov, Instanton like transitions in high-energy collisions. Int. J. Mod. Phys. A **8**, 1823–1886 (1993). https://doi.org/10.1142/S0217751X93000771
29. S.Y. Khlebnikov, V.A. Rubakov, P.G. Tinyakov, Instanton induced cross-sections below the sphaleron. Nucl. Phys. B **350**, 441–473 (1991). https://doi.org/10.1016/0550-3213(91)90267-2
30. A.H. Mueller, On higher order semiclassical corrections to high-energy cross-sections in the one instanton sector. Nucl. Phys. B **364**, 109–126 (1991). https://doi.org/10.1016/0550-3213(91)90580-Q
31. V.V. Khoze, A. Ringwald, Total cross-section for anomalous fermion number violation via dispersion relation. Nucl. Phys. B **355**, 351–368 (1991). https://doi.org/10.1016/0550-3213(91)90118-H
32. D. Diakonov, M.V. Polyakov, Baryon number nonconservation at high-energies and instanton interactions. Nucl. Phys. B **389**, 109–132 (1993). https://doi.org/10.1016/0550-3213(93)90287-Y
33. I. Balitsky, A. Schafer, Valley method versus instanton induced effective Lagrangian up to $(E/E_{sphaleron})^{8/3}$. Nucl. Phys. B **404**, 639–683 (1993). https://doi.org/10.1016/0550-3213(93)90593-E. arXiv:hep-ph/9304261
34. A. Ringwald, Electroweak instantons/sphalerons at VLHC? Phys. Lett. B **555**, 227–237 (2003). https://doi.org/10.1016/S0370-2693(03)00060-1. arXiv:hep-ph/0212099
35. F.L. Bezrukov, D. Levkov, C. Rebbi, V.A. Rubakov, P. Tinyakov, Semiclassical study of baryon and lepton number violation in high-energy electroweak collisions. Phys. Rev. D **68**, 036005 (2003). https://doi.org/10.1103/PhysRevD.68.036005. arXiv:hep-ph/0304180
36. S.H.H. Tye, S.S.C. Wong, Bloch wave function for the periodic sphaleron potential and unsuppressed baryon and lepton number violating processes. Phys. Rev. D **92**, 045005 (2015). https://doi.org/10.1103/PhysRevD.92.045005. arXiv:1505.03690
37. J. Ellis, K. Sakurai, Search for sphalerons in proton–proton collisions. JHEP **04**, 086 (2016). https://doi.org/10.1007/JHEP04(2016)086. arXiv:1601.03654

38. J. Ellis, K. Sakurai, M. Spannowsky, Search for sphalerons: IceCube vs. LHC. JHEP **05**, 085 (2016). https://doi.org/10.1007/JHEP05(2016)085. arXiv:1603.06573
39. G. Brooijmans, P. Schichtel, M. Spannowsky, Cosmic ray air showers from sphalerons. Phys. Lett. B **761**, 213–218 (2016). https://doi.org/10.1016/j.physletb.2016.08.030. arXiv:1602.00647
40. K. Funakubo, K. Fuyuto, E. Senaha, Does a band structure affect sphaleron processes? arXiv:1612.05431
41. S.H.H. Tye, S.S.C. Wong, Baryon number violating scatterings in laboratories. Phys. Rev. D **96**, 093004 (2017). https://doi.org/10.1103/PhysRevD.96.093004. arXiv:1710.07223
42. A. Ringwald, K. Sakurai, B.R. Webber, Limits on electroweak instanton-induced processes with multiple boson production. JHEP **11**, 105 (2018). https://doi.org/10.1007/JHEP11(2018)105. arXiv:1809.10833
43. D.G. Cerdeño, P. Reimitz, K. Sakurai, C. Tamarit, $B + L$ violation at colliders and new physics. JHEP **04**, 076 (2018). https://doi.org/10.1007/JHEP04(2018)076. arXiv:1801.03492
44. A. Papaefstathiou, S. Plätzer, K. Sakurai, On the phenomenology of sphaleron-induced processes at the LHC and beyond. JHEP **12**, 017 (2019). https://doi.org/10.1007/JHEP12(2019)017. arXiv:1910.04761
45. J. Jaeckel, W. Yin, High energy sphalerons for baryogenesis at low temperatures. Phys. Rev. D **107**, 015001 (2023). https://doi.org/10.1103/PhysRevD.107.015001. arXiv:2206.06376
46. M.A. Shifman (ed.), Computation of the quantum effects due to a four-dimensional pseudoparticle. Phys. Rev. D **14**, 3432–3450 (1976). https://doi.org/10.1103/PhysRevD.14.3432. [Erratum: Phys. Rev. D **18**, 2199 (1978)]
47. M.A. Shifman (ed.), Symmetry breaking through Bell–Jackiw anomalies. Phys. Rev. Lett. **37**, 8–11 (1976). https://doi.org/10.1103/PhysRevLett.37.8
48. CMS Collaboration, Search for black holes and sphalerons in high-multiplicity final states in proton–proton collisions at $\sqrt{s} = 13$ TeV. JHEP **11**, 042 (2018). https://doi.org/10.1007/JHEP11(2018)042. arXiv:1805.06013
49. K.S. Thorne, Nonspherical gravitational collapse: a short review
50. P. Schichtel, M. Spannowsky, P. Waite, Constraining strongly coupled new physics from cosmic rays with machine learning techniques. EPL **127**, 61002 (2019). https://doi.org/10.1209/0295-5075/127/61002. arXiv:1906.09064
51. HEP ML Community, A living review of machine learning for particle physics. https://iml-wg.github.io/HEPML-LivingReview/
52. A. Aurisano, A. Radovic, D. Rocco, A. Himmel et al., A convolutional neural network neutrino event classifier. JINST **11**, P09001 (2016). https://doi.org/10.1088/1748-0221/11/09/P09001. arXiv:1604.01444
53. W. Bhimji, S.A. Farrell, T. Kurth, M. Paganini et al., Deep neural networks for physics analysis on low-level whole-detector data at the LHC. J. Phys. Conf. Ser. **1085**, 042034 (2018). https://doi.org/10.1088/1742-6596/1085/4/042034. arXiv:1711.03573
54. M. Andrews, M. Paulini, S. Gleyzer, B. Poczos, Exploring end-to-end deep learning applications for event classification at CMS. EPJ Web Conf. **214**, 06031 (2019). https://doi.org/10.1051/epjconf/201921406031
55. K. Fukushima, Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. **36**, 193–202 (1980). https://doi.org/10.1007/BF00344251
56. G. Veneziano, String-theoretic unitary S-matrix at the threshold of black-hole production. JHEP **11**, 001 (2004). https://doi.org/10.1088/1126-6708/2004/11/001. arXiv:hep-th/0410166
57. D.-C. Dai, G. Starkman, D. Stojkovic, C. Issever et al., BlackMax: a black-hole event generator with rotation, recoil, split branes, and brane tension. Phys. Rev. D **77**, 076007 (2008). https://doi.org/10.1103/PhysRevD.77.076007. arXiv:0711.3012
58. M. Bahr et al., Herwig++ physics and manual. Eur. Phys. J. C **58**, 639–707 (2008). https://doi.org/10.1140/epjc/s10052-008-0798-9. arXiv:0803.0883
59. S. Gieseke et al., Herwig++ 2.5 release note. arXiv:1102.1672
60. K. Arnold et al., Herwig++ 2.6 release note. arXiv:1205.4902
61. J. Bellm et al., Herwig++ 2.7 release note. arXiv:1310.6877
62. J. Bellm et al., Herwig 7.0/Herwig++ 3.0 release note. Eur. Phys. J. C **76**, 196 (2016). https://doi.org/10.1140/epjc/s10052-016-4018-8. arXiv:1512.01178
63. J. Bellm et al., Herwig 7.1 release note. arXiv:1705.06919
64. J. Bellm et al., Herwig 7.2 release note. Eur. Phys. J. C **80**, 452 (2020). https://doi.org/10.1140/epjc/s10052-020-8011-x. arXiv:1912.06509
65. A. Papaefstathiou, S. Plätzer, K. Sakurai, On the phenomenology of sphaleron-induced processes at the LHC and beyond. J. High Energy Phys. **2019**, 17 (2019). https://doi.org/10.1007/JHEP12(2019)017. arXiv:1910.04761
66. DELPHES 3 Collaboration, DELPHES 3, a modular framework for fast simulation of a generic collider experiment. JHEP **02**, 057 (2014). https://doi.org/10.1007/JHEP02(2014)057. arXiv:1307.6346
67. T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16* (Association for Computing Machinery, New York, 2016), pp. 785–794. https://doi.org/10.1145/2939672.2939785
68. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. arXiv e-prints (2015). arXiv:1512.03385
69. A. Grefsrud, choisant/imcalML: 1.0 (2023)
70. C. Adam-Bourdarios, G. Cowan, C. Germain-Renaud, I. Guyon et al., The Higgs machine learning challenge. J. Phys.: Conf. Ser. **664**, 072015 (2015). https://doi.org/10.1088/1742-6596/664/7/072015
71. ATLAS Collaboration, Search for dark matter in events with missing transverse momentum and a Higgs boson decaying into two photons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. JHEP **10**, 013 (2021). https://doi.org/10.1007/JHEP10(2021)013. arXiv:2104.13240
72. A. Paszke, S. Gross, F. Massa, A. Lerer et al., PyTorch an imperative style, high-performance deep learning library. CoRR (2019). arXiv:1912.01703
73. L.N. Smith, Cyclical learning rates for training neural networks (2017). arXiv:1506.01186
74. D. P. Kingma, J. Ba, Adam: a method for stochastic optimization, in *International Conference on Learning Representations* (2014). arXiv:1412.6980