



# Model-independent searches of new physics in DARWIN with deep learning

XLZD Collaboration\*

Received: 21 October 2024 / Accepted: 3 December 2025  
© The Author(s) 2026

**Abstract** We present a deep learning pipeline to perform a model-independent, likelihood-free search for anomalous (i.e., non-background) events in the proposed next-generation multi-ton scale liquid xenon-based direct detection experiment, DARWIN. We train an anomaly detector comprising a variational autoencoder (VAE) and a classifier on high-dimensional simulated detector response data and construct a 1D anomaly score to reject the background-only hypothesis in the presence of an excess of non-background-like events. We use simulated validation data to determine the power of the method to reject the background-only hypothesis in the presence of a WIMP dark matter signal, without any model-dependent assumption about the nature of the signal. We show that our neural networks learn relevant features of the events from low-level, high-dimensional detector outputs, avoiding lossy and computationally expensive compression into lower-dimensional observables. Our approach is complementary to the usual likelihood-based analysis, in that it reduces the reliance on many of the corrections and cuts that are traditionally part of the analysis chain, with the potential of achieving higher accuracy and significant reduction of analysis time. We envisage the methodology presented in this work augmenting or complementing likelihood-based and other data-driven methods currently utilized in the DARWIN (and in the future, XLZD) analysis pipeline.

## 1 Introduction

A promising method for investigations of the ever-elusive dark matter sector involves seeking excess nuclear recoils in subterranean detectors, a strategy known as direct detection (DD) [1]. Over the years, a number of xenon (XENONnT [2], LUX-ZEPLIN (LZ) [3], PandaX[4]) and argon (DEAP-3600 [5], DarkSide-20k [6], ArDM [7]) ton-scale experiments have steadily enhanced the sensitivity to physics beyond the standard model (BSM), and this effort is expected to continue, with plans for a next-generation dark matter

and neutrino observatory. While earlier designs for a ‘dark matter WIMP search with liquid xenon’ observatory (DARWIN) [8,9] aimed at an active liquid xenon target mass of 40 tons, the recently formed XLZD Collaboration proposes an even more ambitious target mass of 60–80 tons [10]. While the design of the XLZD experiment is being developed, this paper focuses on DARWIN, a well-defined proposal for a large-scale observatory using a xenon dual-phase time projection chamber (TPC) to study phenomena requiring low-background conditions. DARWIN aims to be sensitive to weakly interacting massive particle (WIMP) dark matter as well as neutrinoless double beta decay, axion-like particles, and any other BSM particles that would manifest through significant interaction with a xenon target. The aim of this work is to introduce a signal model-agnostic, deep learning-based analysis pipeline, offering a complementary and alternative approach to the standard likelihood-based analysis chain in such a detector. The benefits of this approach are that it enables a fuller exploitation of the detector readout data, without the information loss potentially incurred in using only hand-crafted summary statistics (such as cS1 and cS2, the corrected prompt primary scintillation and secondary electroluminescence of ionized electrons signals, respectively), and that it can include in the pipeline any physics effect that can be faithfully simulated, including systematics.

Machine learning (ML) has emerged as a powerful tool within the physics community, and its relevance to DM phenomenology has been growing rapidly [11–15].

Unsupervised machine learning has been increasingly employed in collider physics to identify anomalies in data, as demonstrated in several recent studies [16–27], with early example applications on simulated events of CMS and ATLAS already in Refs. [28,29], as well as Ref. [30], where an “anomaly awareness” algorithm is proposed. ML techniques were also applied to DD experiments for a variety of tasks, ranging from signal classification to fast likelihood evaluation [31–36]. Ref. [32] utilizes a semi-supervised deep neural network comprising a pretrained convolutional

\* e-mail: [ascafid@sissa.it](mailto:ascafid@sissa.it)

neural network (CNN) and a VAE in order to detect the presence of excess nuclear recoils above the expected background in DD experiments.

The established approach to the detection of a new physics signal in DD experiment with dual dual-phase target is a likelihood-based test with an assumed asymptotic distribution [9], with the likelihood a function of the so-called “corrected” S1 and S2 signals (cS1 and cS2, respectively). By using neural networks that are trained on high-dimensional representations of detector events, we show in this paper that it is possible to infer the relevant properties (energy distribution, type of recoil) from detector-level readouts, without the approximation and loss of information incurred in the usual cS1, cS2 compression. This opens the door to the possibility of an end-to-end inference approach that is fully simulation-based, including all necessary corrections and cuts that are traditionally done in the analysis and inference chain, a process which takes up a significant fraction of analysis time in current-generation detectors. This approach relies however, on the availability of accurate and faithful simulations: real detectors and backgrounds are usually more complex and/or feature unexpected characteristics that deviate from simulations. Data-driven calibration and adversarial training techniques can help mitigate such systematic differences, improving robustness against these biases – something we plan to explore in future works.

Subject to the above caveat, the aim of this paper is to demonstrate the capability of a deep learning pipeline to detect the presence of an ‘anomalous’ signal above a known (from simulations) background in DARWIN, without explicit modeling of the likelihood nor of the physics underlying the anomaly (i.e., without assuming a specific dark matter model). In this sense, our analysis is model-independent, that is, agnostic to any specific new physics model. We achieve this by training an anomaly detector on event-by-event simulated detector response quanta using the DARWIN simulation pipeline, and by constructing an anomaly score designed to maximize the sensitivity to rejecting the background-only hypothesis. The choice of DARWIN as a case study is motivated by the availability of sufficiently mature and complete detector simulations, which is not yet the case for XLZD. Of course, the general approach is applicable to future detectors, once their design and simulation pipeline are settled. Application of this approach to existing detectors would require refinement to account for rare and/or unforeseen backgrounds or detector effects that may not be simulated correctly. Since this paper focuses on demonstrating the overall methodology, we leave exploration of such issues to future investigations.

This paper is structured as follows. In Sect. 2 we briefly introduce the design of the DARWIN detector, we describe the data structure used to train the model, as well as the simulations that were employed to this end. In Sect. 3 we explain

the aim of the analysis, the methodology employed and its novelty. We also present the detailed simulation pipeline adopted for the study, the split between training and validation sets and the training procedure. In Sect. 4, we validate our approach by determining the sensitivity of DARWIN to rejecting the background-only null-hypothesis in the presence of a simulated injection of a WIMP signal. We then conclude in Sect. 5.

## 2 Experiment design and data simulation

### 2.1 The DARWIN detector design

DARWIN is conceived as a multi-ton, dual-phase liquid xenon time-projection chamber (TPC) designed to push DD sensitivity to the verge of the astrophysical neutrino floor [37]. The reference design holds  $\sim 50$  t of xenon, with about 40 t active, inside a  $2.6 \text{ m} \times 2.6 \text{ m}$  cylindrical TPC; prompt VUV scintillation (S1) and proportional electroluminescence (S2) are captured by matched top and bottom arrays of ultra-low-background photomultipliers (PMTs) or silicon photomultiplier (SiPM) tiles, providing sub-keV thresholds and event-by-event electron vs nuclear recoil discrimination. The large homogeneous target, excellent self-shielding and simultaneous light-and-charge readout make large TPC chambers versatile platforms for dark matter, neutrino and rare decay physics [8].

The TPC design is suspended in a double-walled low-radioactivity cryostat and immersed in an instrumented water tank that serves both as a passive  $\gamma/n$  shield and an active Cherenkov muon veto. A uniform drift field of the order of  $0.5 \text{ kV cm}^{-1}$  is generated inside the TPC, enabling electrons to traverse the full 2.6 m height. This long-drift capability- as well as cryogenics, purification, and DAQ concepts, has been validated in the Xenoscope vertical demonstrator and related optical simulation test-stands [38], as well as a second large scale demonstrator called PANCAKE [39].

In 2024 the DARWIN, LZ and XENONnT collaborations unified their efforts in the next-generation XLZD programme [10], which scales the dual-phase concept to 60–80 t of active xenon while retaining the core detector architecture. DARWIN’s hardware prototypes and simulation tools remain the principal testbeds for XLZD component development and the waveform-level analysis showcased here. Consequently, the study performed in this paper adopts the original 40 t DARWIN geometry when generating simulated S1/S2 events, with the ML methodology and data analysis pipeline having direct application to any future XLZD-type detector.

## 2.2 Generation of simulated events

Our simulation-based pipeline is reliant on the quality of the simulations adopted. For this reason, we use state-of-the-art simulations tailored to the DARWIN design. We use the Geant4 transport code [40] within the DARWIN-Geant4 framework [41] to handle the tracking of particles within a rendering of the detector geometry. The Noble Element Simulation Technique (NEST) v2.3.12 [42] handles the microphysics of how particles interact with the active xenon volume. NEST provides a robust and well-established framework that simulates the atomic and nuclear physics involved in energy deposition and the corresponding response of the detector, and generates the light and charge yields for each type of interaction within the detector. These simulated light and charge yields are compared and calibrated against previous xenon experiments, see Ref. [9] for details. Full signal propagation and observable readout within the TPC volume that produced the simulated waveforms and PMT hit-patterns were produced by custom-written detector simulation code based on the Tray [43] architecture.

Any WIMP search relies on distinguishing between background events and the WIMP-induced signal. We therefore need our deep learning pipeline to learn to characterize the background distribution. The majority of background at DARWIN will be electron recoil (ER) events originating from various terrestrial and cosmogenic sources, while nuclear recoil (NR) backgrounds remain in the form of irreducible cosmogenic neutrinos and sub-dominant radiogenic neutrons [41, 44], which must be included as part of the background simulation. WIMPs of mass  $\mathcal{O}( > 1 )$  GeV deposit their energy into the detector via NR events.

We describe the background simulations used in this study in Sect. Appendix A, and give here only a concise summary. For each type of background (ER and NR), events with uniformly distributed recoil energies were simulated in the range 1–100 keV. The simulations include detector response effects (including electron-ion recombination, electron drift, and photon-collection efficiency), which transform the raw energy deposition from the initial particle interaction into the observable signals in the detector.<sup>1</sup>

For our analysis, we follow the approach taken in Ref. [32], and adopt as description of the TPC data the total S1 + S2 waveforms (i.e, signal as a function of time, summed over all individual PMTs), as well as the top and bottom S2 PMT hit pattern readout.<sup>2</sup> We use the total waveforms (as opposed to

<sup>1</sup> Another form of background observed by XENONnT was anomalous events emanating from radioactivity in the TPC walls, referred to as ‘surface’ events. We did not simulate such event realizations in this study, but work is being conducted to implement them with unsupervised veto models.

<sup>2</sup> We note that not including the S1 top and bottom PMT hit-patterns decreases sensitivity to so-called ‘ $\gamma$ -X’ and ‘neutron-X’ events

the PMT-specific waveform) in order to reduce the dimensionality and complexity of the data vector provided to the neural networks. To exploit the detector readout data in even more fundamental form, one should adopt a model capable of learning a representation of the PMT responses from the entire PMT array in the temporal domain [46, 47] – something the method in this work is unable to scale to. Modern developments in Transformer or graph neural network architectures could potentially be used for handling time-domain individual PMT readouts [48–50]. In order to meet this challenge however, we plan to utilize the Rotary Masked Autoencoder of Ref. [51].

In Fig. 1 we show an example of the data used to train the neural networks. Events are simulated in a fiducial detector volume (FV) of 31.5 t, chosen to optimize the detection of rare NR while minimizing ER background interference towards the boundaries of the bulk xenon, as well as other factors [9]. The simulations are realized with a drift field of 200.0 V/cm, registering events when at least 4 photons are detected within a 200-nanosecond window (referred to as a ‘4-fold coincidence’, or N4T200). We do not utilize spatial reconstruction to provide a further fiducialization cut. Work is being done in this direction at XENON, see for example Ref. [52].

## 3 Methodology

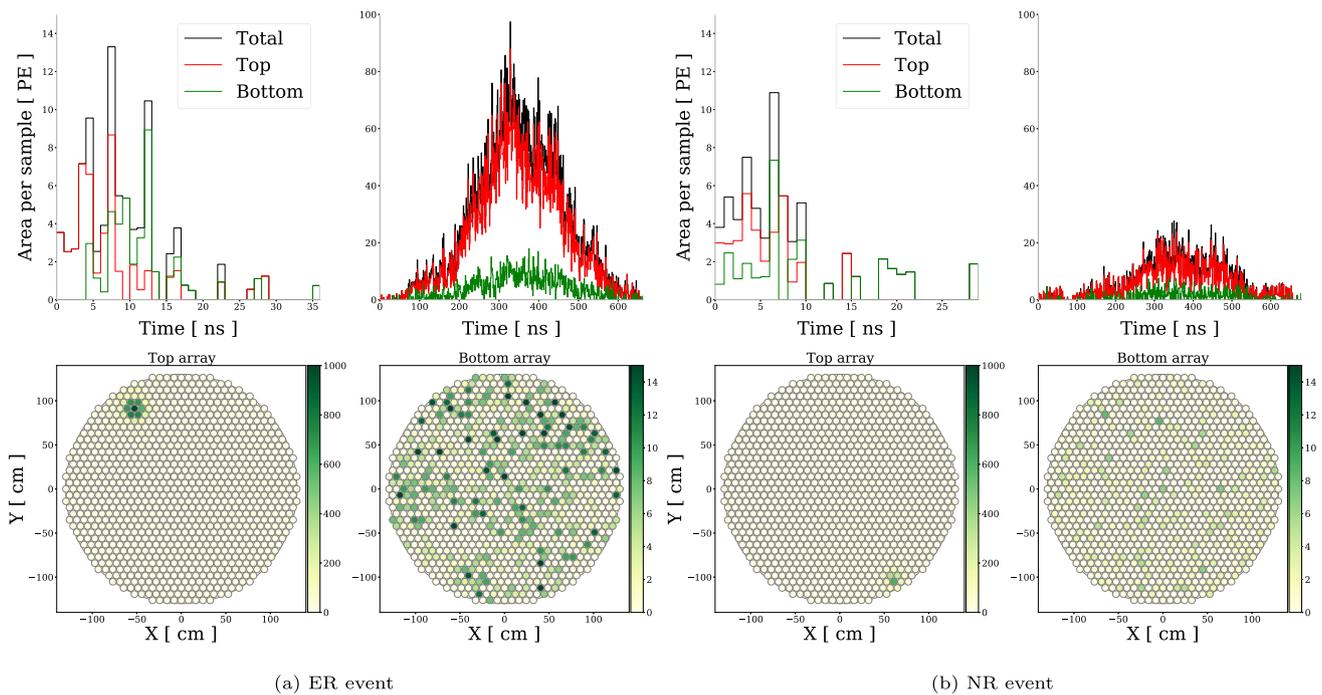
In this section, we first provide an overview of the objective of this study, followed by a concise description of the analysis methodology, which highlights the novelty of the approach. The architectural details as well as hyperparameters of the VAE and classifier used in this study are detailed in Appendix B.

### 3.1 Simulation-based anomaly detection

The objective of this study is to demonstrate the potential of a deep learning pipeline to detect a WIMP-like signal above known simulated backgrounds in a semi-supervised fashion. This is complementary to the traditional likelihood-based method, as it offers several potential advantages: first, our approach makes fuller use of the information contained in the PMT readout data, thus avoiding the information loss that compression into summary statistics (such as cS1/cS2) inevitably incurs; secondly, it can incorporate in the pipeline any effect that can be faithfully simulated in the mock data.

Footnote 2 continued

as observed by XENON100 [45, 46]. Since we do not include such background events in this study, this has no bearing on our results here. Future developments will add both ‘ $\gamma$ -X’ and ‘neutron-X’ events to the background, and S1 PMT patterns to the input data.



**Fig. 1** Example of simulated detector observables of an electron recoil (ER) **a** and nuclear recoil (NR) **b** event in DARWIN. **Top:** Number of S1 (left sub-panel) and S2 (right sub-panel) photoelectrons (PE) as a function of time after initial S1 triggering. Red (green) denotes obser-

vation in the top (bottom) PMT array. The black curves are the total S1 + S2 and are used for training the neural networks. **Bottom:** Top and bottom S2 PMT deposit spatial pattern. The color bar indicates the PMT hit count. These data are used to train the neural networks

This means that the impact of nuisance parameters can be accounted for by simply including their sampling within the generation of training data. Finally, our approach does not rely on approximations to the likelihood, nor to a model-specific form of the WIMP-signal, therefore being more general and model-agnostic.

Our aim is to train a suitable neural network to identify anomalous signals – i.e., any event that can be distinguished statistically from the simulated ER and NR background distribution. This involves the computation of an ‘anomaly score’,  $TS$ , obtained from the combined loss distribution and classification output of a neural anomaly detector. The anomaly score is used to ascertain whether a collection of observed events  $\mathbf{X}_n = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ , deviates from the background-only distribution. The null hypothesis, which we denote  $\mathcal{H}_0$ , is that the events  $\mathbf{X}_n$  are drawn from a distribution where no signal is present, i.e., compatible with the expected background.

The anomaly detector consists of two parts: a supervised binary classifier and a VAE. The classifier learns from training data to distinguish ER from NR events, whilst the VAE is trained solely on ER events<sup>3</sup>. After training, validation

data (i.e., that the network has not been trained on) is given to the network, and its  $TS$  distribution obtained: events that deviated from background-like properties will manifest in the 1D space of the  $TS$  distribution as an excess over the background-only distribution. A simple 1D statistical test is then employed to reject the background-only hypothesis.

### 3.2 Definition and distribution of the anomaly score

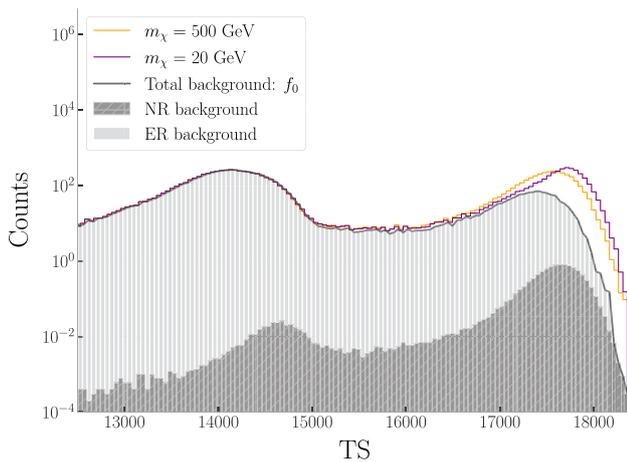
The anomaly score,  $TS$ , is defined as the weighted linear combination of the reconstruction loss from the VAE, or ‘ELBO’ (see Eq. B.2), and the classifier’s binary cross-entropy,  $H_B$ , so that larger values correspond to deviations from the null hypothesis:

$$TS = (-\text{ELBO}) + RH_B \\ = D_{\text{KL}}(q(\mathbf{z}|\mathbf{x}_{\text{in}})||p(\mathbf{z})) - \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_{\text{in}})}[\log p_{\mathbf{x}_{\text{in}}}(\mathbf{x}_{\text{D}}|\mathbf{z})] \quad (1)$$

Footnote 3 continued

representative sample of *all* event classes (comprising both ER and NR) as well as calibration data. This allows potentially anomalous events to be identified in the latent space of the autoencoder. Representation learning of this type can be useful in that it is one-sample and ‘data-driven’, at the expense of sensitivity to an explicit null-hypothesis.

<sup>3</sup> An alternative approach (see, for example, the work of the LUX-ZEPLIN (LZ) collaboration in Ref. [53]) can be to train a VAE on a



**Fig. 2** Distribution of the anomaly score  $TS$  from a pseudo-dataset used in this study. The stacked gray bars represent the  $TS$  distribution for the ER (light gray) and NR (dark gray) background. The colored lines are the distributions in  $TS$  after the injection of signal components for 20 and 500 GeV WIMPs, with a scattering cross-section of  $\sigma_\chi = 10^{-46} \text{ cm}^2$  (a large value chosen for clarity of illustration). The binning is illustrative, as our sensitivity analysis is unbinned. The solid black line is the total background pdf  $f_0$

$$\begin{aligned}
 &+ R H_B(\mathbf{x}_{in}) \\
 &= -\frac{1}{2}\beta \sum_{j=1}^m \left(1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2\right) \\
 &- \log \mathcal{N}_{\mathbf{x}_{in}}(\mathbf{x}_D, \text{diag}(\sigma_D)^2) - R \log(1 - p(\mathbf{x}_{in})) . \quad (2)
 \end{aligned}$$

The hyperparameter  $R$  controls the relative importance of the binary cross-entropy term, and its optimization is discussed in Appendix C.

In order to determine the  $TS$  distribution under  $\mathcal{H}_0$ , a set of  $10^4$  ER and  $10^4$  NR events are simulated according to their expected rates after trigger-level cuts, fiducialisation and signal region cuts, as given in Fig. 5 of Appendix A. In Fig. 2 we show a dataset comprised of each background component (dark/light grey histogram) as well as two injected WIMP signals (color curves) at a relatively large cross-section (for illustration purposes) in  $TS$  space, re-weighted to an exposure of 200 ty. The spectral dependence of the ELBO manifests in  $TS$  space, with anomalous events (in this case, WIMPs) being mapped to larger  $TS$  values than the background. We therefore observe two bumps in the  $TS$  distribution of the NR and ER backgrounds corresponding to the classifier’s prediction. ER events that present with higher  $TS$  values typically have lower energies, as would make qualitative sense due to low-energy ER being indistinguishable from NR. In Appendix D, we demonstrate that the VAE non-trivially encodes the spectral energy information of all events (both NR and ER), despite the VAE having been trained only on ER events.

### 3.3 Neural networks training and validation

The neural networks are trained on vectorized formats: [S1WaveformTotal, S2WaveformTotal, S2Patterns], with a total size of 3835. The waveform and hit pattern data provide information about each event, making it possible for the neural anomaly detector to learn complex features pertaining to the class of the event (ER vs NR) as well as the different spectral dependency of each class (see Appendix D and Appendix E for further details).

We generate training data sets consisting of an even sample of  $2 \times 10^4$  ER and NR events with true recoil energies uniformly distributed in  $E_R \in [1, 100]$  keV, with 30% being kept aside for validation. The average training time per epoch is  $\sim 1$  s for the VAE ( $\sim 40$  seconds total training time) and  $\sim 0.8$  seconds for the classifier ( $\sim 8$  seconds total training time) on an NVIDIA A100-PCIE-40GB GPU. Testing times event-by-event are of the order of ms.

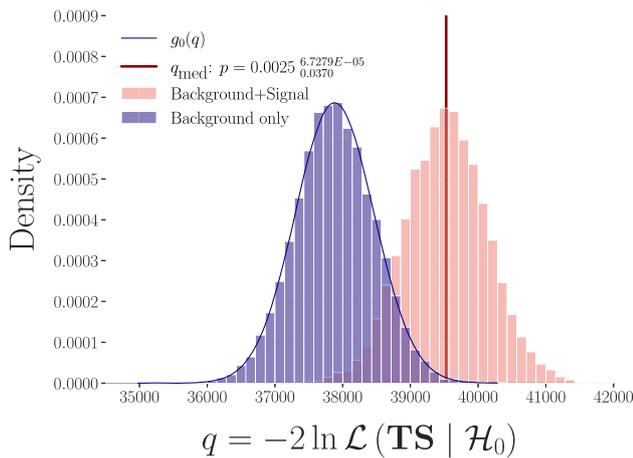
### 3.4 Null hypothesis test

In order to test for the presence of an anomalous bump (due to anomalous, non-background-like events) in the  $TS$  distribution, we define an unbinned 1D likelihood for the background probability distribution function (pdf),  $f_0$ , called the ‘extended Poisson’ [54]:

$$\mathcal{L}(\mathbf{TS}|\mathcal{H}_0) = \frac{e^{-B}}{N!} \prod_{i=1}^N B f_0(TS_i) . \quad (3)$$

Here  $\mathbf{TS}$  denotes the vector of observed  $TS$  produced by the trained neural network for events labeled by  $i$  during a given exposure, while  $B$  is the total expected number of background events and  $N$  is the number of observed events.

We take as a test statistic the distribution of  $q = -2 \ln \mathcal{L}$ , formalizing  $\mathcal{H}_0$  as the asymptotic distribution of  $q$  after simulating  $\sim 10^4$  experiments, each with an exposure of 200 ty, using pseudo-datasets comprised solely of background events, where the number of events per experiment is sampled from a Poisson with expectation value  $B$ , leading to a number of events per experiment  $\sim \mathcal{O}(6.5 \times 10^3)$ . This distribution of  $q$  is shown in blue in Fig. 3. Any upward fluctuation of the negative log-likelihood denotes a departure from the background-only hypothesis by construction. The distribution of  $q$  from another  $10^4$  simulated experiments including an injected WIMP signal at a fixed benchmark of  $\sigma = 6.5 \times 10^{-48} \text{ cm}^2, m_\chi = 50 \text{ GeV}$  is shown in pink, while the median significance  $q_{\text{med}}$  (i.e., the median  $p$ -value for which one can reject  $\mathcal{H}_0$  in the presence of a signal, calculated over a collection of pseudo-datasets [55]) is denoted by the vertical red line. The median sensitivity is the  $p$ -value



**Fig. 3** Distribution of  $q = -2 \ln \mathcal{L}(\text{TS} | \mathcal{H}_0)$  from pseudodata generated under  $\mathcal{H}_0$  (blue) and with an injected dark matter (WIMP) signal with  $\sigma_{\text{SI}} = 6.5 \times 10^{-48} \text{ cm}^2$  and  $m_\chi = 50 \text{ GeV}$  (pink), which yields a median sensitivity of  $\sim 3\sigma$  at 200ty exposure. We also display as a blue line the kernel density estimate (KDE) used to evaluate the integral in Eq. (4). The red vertical line denotes  $q_{\text{med}}$

to reject  $\mathcal{H}_0$  corresponding to  $q_{\text{med}}$ :

$$p_{\text{med}} = \int_{q_{\text{med}}}^{\infty} dq g_0(q), \quad (4)$$

where  $g_0(q)$  is the distribution of  $q$  under the null hypothesis.

## 4 Results

In this section, we present the results from our approach on simulated data. For this analysis, the ER and NR background distributions have been re-weighted to their expected values using the background benchmarks from Appendix A. The median sensitivity to reject  $\mathcal{H}_0$  as a function of exposure is shown as the red line in Fig. 4 (left panel) for the WIMP benchmark adopted in Fig. 3 ( $\sigma_{\text{SI}} = 6.5 \times 10^{-48} \text{ cm}^2$ ,  $m_\chi = 50 \text{ GeV}$ ).

The right panel of Fig. 4 shows the median sensitivity in the canonical 2D WIMP parameter space for a fixed exposure of 200 ty. We plot the median sensitivity as a color gradient, indicating contours corresponding to 1, 2 and  $3\sigma$  median sensitivity. For qualitative comparison only, we display the 2016 median DARWIN 90% C.L. upper limit sensitivity as a green dashed curve [8]. It is important to note that this 90% C.L. upper limit sensitivity is not directly comparable to the background rejection test in our pipeline, as these are two fundamentally different statistical tests: the 90% C.L. upper limit sensitivity is model-dependent (as the WIMP signal is specific for a given model), whilst the anomaly detection method is agnostic to the WIMP physics, as the neural networks were only trained on samples indicative of a

background-only dataset, with no information about WIMP-like events. Hence, whilst the background rejection  $p$ -value we present is a somewhat ‘stronger’ statistical claim (in that it is model-independent), we find (as expected) that an upper limit in the presence of an explicit alternative WIMP model is significantly more constraining.

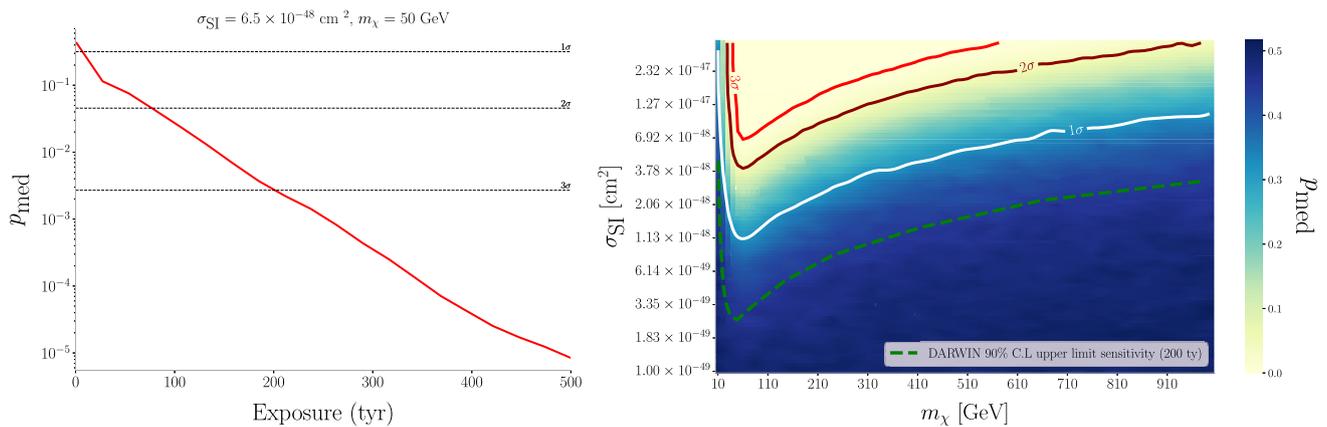
## 5 Conclusions

This study presents the foundation for a deep learning analysis pipeline to perform anomaly detection in next next-generation dark matter direction detection experiment – in this case, the DARWIN design. The proposed methodology provides a prototype for future developments in statistical inference in rare physics searches with xenon-based TPCs, and promises to extract maximal information from the high-dimensional event data produced by TPC experiments. This is particularly critical given the current challenges faced by modern TPC experiments, where a substantial portion of analysis time is devoted to tuning optimal cuts and corrections for high-level, compressed summary observables.

The method in this paper presents an anomaly-aware machine learning technique that leverages deep learning to conduct a background rejection task. We use a neural network architecture consisting of an unsupervised VAE and a fully connected classifier that extracts relevant event-by-event features (including energy information) from PMT hit pattern data and total S1 and S2 waveforms. We find that the neural anomaly detector achieves sensitivity to reject  $\mathcal{H}_0$  at the order of  $3\sigma$  after  $\sim 200 \text{ ty}$  for a WIMP benchmark of  $\sigma_{\text{SI}} = 6.5 \times 10^{-48} \text{ cm}^2$ ,  $m_\chi = 50 \text{ GeV}$ .

A model-independent anomaly detection can serve as a ‘first pass’ analysis, assessing if there is any data that is not consistent with the background-only expectation, before moving on to a more sensitive model-dependent search (e.g., via likelihood ratio). Whilst we have validated our pipeline in the context of a canonically interacting WIMP, the machinery remains identical for any new physics search. This makes the development and deployment of these types of analyses an important addition to the standard statistical pipeline.

As is always the case for simulation-based analyses, the neural networks could be subject to missing or misinterpreting key underlying features or stochastically of real data should the simulations be incomplete or otherwise imperfect [56,57]. To mitigate this risk, one could expand the pipeline to include fine-tuning the models on calibration data in the training of the neural network, thereby complementing simulated events with actual observations. A large computational effort is currently being directed toward folding in calibration information into the derivation of the high-level cS1/cS2 statistics, something that would be complemented by our approach: a neural network-based analysis pipeline can alle-



**Fig. 4** **Left:** Median sensitivity to reject the background-only hypothesis as a function of detector exposure at the benchmark  $\sigma_{\text{SI}} = 6.5 \times 10^{-48} \text{ cm}^2$ ,  $m_\chi = 50 \text{ GeV}$ . Thresholds of 1,2 and 3 $\sigma$  decision boundaries are shown as black horizontal dashed lines. **Right:** Median

sensitivity in the  $m_\chi$ ,  $\sigma_{\text{SI}}$  plane from the anomaly detection pipeline (exposure of 200 ty), with contours at 1, 2 and 3 $\sigma$  (solid lines). For qualitative comparison, the WIMP-model dependent DARWIN 90% C.L. median upper limit sensitivity is shown as the green dashed line

violate the computational burden as it bypasses the need for these corrections. However, care must be taken with uncertainties due to specification of the recoil energy of events, especially at lower energy thresholds [58,59]. This type of issue could be circumvented with unsupervised anomaly detector networks that have integrated domain adaptation between simulated source data and target calibration [60]. Investigation of these types of models will be the subject of future work.

Given the simulation-rich environment at DARWIN and in the future, XLZD, we plan to leverage this approach, including multi-scatter classification, energy and position reconstruction, circumventing the need for traditional detector fiducialisation or signal region definition. Other architecture developments will be aimed at handling high-dimensional temporal PMT data, accidental coincidence, and surface events background discrimination, as well as inter-ER background classification.

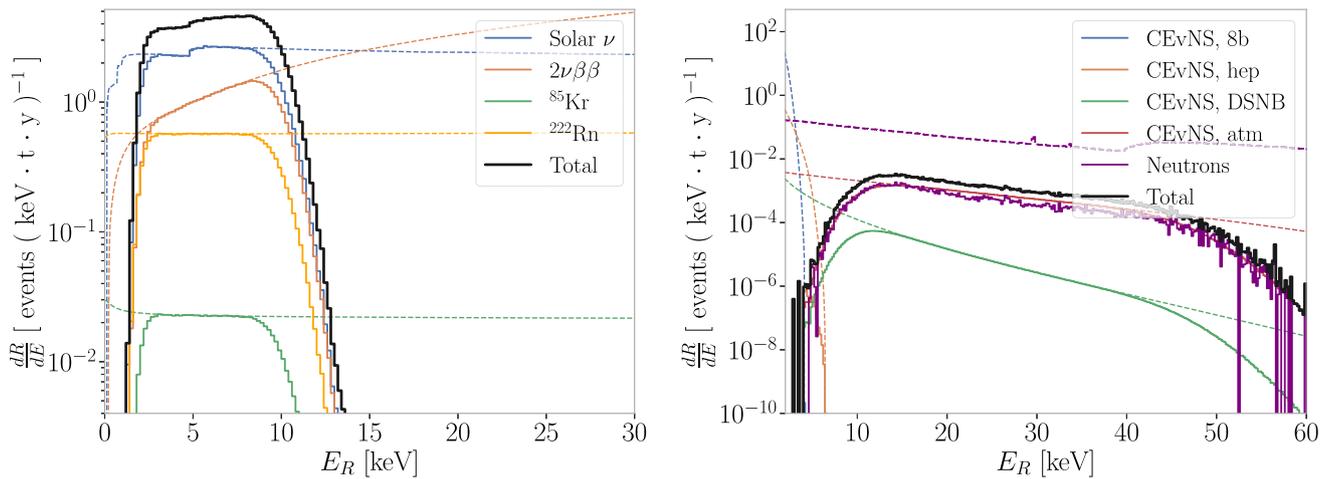
**Funding** AS was partially supported by the grant “DS4ASTRO: Data Science methods for Multi-Messenger Astrophysics & Multi-Survey Cosmology”, in the framework of the PRO3 ‘Programma Congiunto’ (DM n. 289/2021) of the Italian Ministry for University and Research. RT and AS acknowledge funding from Next Generation EU, in the context of the National Recovery and Resilience Plan, Investment PE1-Project FAIR “Future Artificial Intelligence Research”. This resource was co-financed by the Next Generation EU [DM 1555 del 11.10.22]. RT is partially supported by the Fondazione ICSC, Spoke 3 “Astrophysics and Cosmos Observations”, Piano Nazionale di Ripresa e Resilienza Project ID CN0000013 “Italian Research Center on High-Performance Computing, Big Data and Quantum Computing” funded by MUR Missione 4 Componente 2 Investimento 1.4: Potenziamento strutture di ricerca e creazione di “campioni nazionali di R&S (M4C2-19)” - Next Generation EU (NGEU). This work was also supported by the Swiss National Science Foundation under grants No 200020-162501 and No 200020-175863, by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agree-

ments No 674896, No 690575 and No 691164, by the European Research Council (ERC) grant agreements No 742789 (Xenoscope) and No 724320 (ULTIMATE), by the Max-Planck-Gesellschaft, by the Deutsche Forschungsgemeinschaft (DFG) under GRK-2149, by the US National Science Foundation (NSF) grants No 1719271 and No 1940209, by the Dutch Science Council (NWO), by the Portuguese FCT, by the Ministry of Education, Science and Technological Development of the Republic of Serbia and by grant ST/N000838/1 from Science and Technology Facilities Council (UK). We further acknowledge funding from the German Federal Ministry for Research, Technology and Space (BMFT) and from the Helmholtz Association.

**Data Availability Statement** This manuscript has no associated code/software. [Author’s comment: Data sharing not applicable to this article as no datasets were generated or analysed during the current study.]

**Code Availability Statement** This manuscript has no associated data. [Author’s comment: Code/Software sharing not applicable to this article as no code/software was generated or analysed during the current study.]

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.  
Funded by SCOAP<sup>3</sup>.



**Fig. 5** Benchmark DARWIN background differential recoil rate spectra considered in this analysis, before (dashed lines) and after (solid lines) detector-level SR, fiducialization and threshold cuts. The total background contributions are shown by black solid lines. **Left:** ER backgrounds originating from low-energy solar neutrinos, two-neutrino double-beta decays of  $^{136}\text{Xe}$  and intrinsic backgrounds from  $^{85}\text{Kr}$  and

$^{222}\text{Rn}$ . **Right:** NR background contributions, produced by coherent neutrino-nucleus scattering sources: solar neutrinos originating from  $^8\text{B}$  and from the helium-proton reaction, atmospheric neutrinos, the diffuse supernova neutrino background, and radiogenic neutrons from the detector. The mean integrated rates are given in Table 1 for each background component

**Table 1** Summary of mean background event rates after detector-level SR, fiducialization and threshold cuts

Background event rates	
	Rate [(ty) $^{-1}$ ]
ER intrinsic: $^{136}\text{Xe}$ ( $2\nu\beta\beta$ )	9.4
ER intrinsic: $^{222}\text{Rn}$	4.5
ER intrinsic: $^{85}\text{Kr}$	0.18
ER solar neutrinos	20.0
NR solar CEvNS	$5.3 \times 10^{-4}$
NR atmospheric CEvNS	$2.6 \times 10^{-2}$
NR radiogenic neutrons	$2.2 \times 10^{-3}$

## Appendix A Background modeling

In this section, we detail the background modeling of this study. The different sources of ER and NR backgrounds relevant to DARWIN are described in Refs. [8, 41, 44].

The background contributions in DARWIN can be categorized into external and intrinsic backgrounds: external backgrounds include gamma-rays and neutrons originating from radioactive decays or interactions outside of the target volume. These can be significantly reduced by target fiducialization due to the high density of liquid xenon. Intrinsic backgrounds, on the other hand, are uniformly distributed in the target region and cannot be reduced by fiducialization.<sup>4</sup>

<sup>4</sup> In this study, we neglect surface events [61] and isolated light and charge signals from accidental coincidences [9] that were considered

The background is obtained after detector-level cuts, including the finite energy threshold of the detector, the fiducial region and signal region (SR) cuts on the combined energy scale (CES), and an estimate of the true deposited recoil energy,  $E_R$ . For this analysis, we adopt a standard value of 31.5 t [8], using an estimated location in the detector for fiducialization cuts. Furthermore, given that the spectral information of all relevant backgrounds is not currently fully known, we apply a [2–10] keVee cut on the CES of each event in line with previous studies [44]. This leaves ERs with a ground truth  $E_R$  between  $\sim$ [2–14] keV and NRs between  $\sim$ [2–60] keV, which are displayed in Fig. 5. A further assumption we make is that multi-scatter events are fully vetoed. Thus, the analysis presented in this work assumes 100% single-scatter selection efficiency.<sup>5</sup> Each background contribution has an expected rate as shown in Table 1.

**ER backgrounds:** Solar neutrinos produced through the proton–proton ( $pp$ ) fusion process and the subsequent beryllium-7 ( $^7\text{Be}$ ) reaction in the Sun are the dominant source of ER background for dark matter searches beyond the ton-scale. This is because of their relatively low energies and high abundance, along with the fact that their contribution cannot be reduced by target purification, fiducialization, nor single-scatter selection. Intrinsic backgrounds, including contribu-

Footnote 4 continued

in the analyses of XENONnT and LZ. Modeling these backgrounds is under current development at DARWIN/XLZD, and so we leave their treatment to future work.

<sup>5</sup> Work is being conducted to incorporate multi-scatter selection using deep learning to supplement the pipeline presented in this work.

tions from isotopes such as  $^{85}\text{Kr}$ , a beta-emitter present in natural krypton, and  $^{222}\text{Rn}$ , are included. These intrinsic backgrounds are uniformly distributed in the target due to the chemical inertness of noble gases. Two-neutrino double-beta decays ( $2\nu\beta\beta$ ) of  $^{136}\text{Xe}$  yield a background that steeply rises with recoil energy. Finally, ER backgrounds originating from  $\gamma$ -rays from radioactive contamination in the cryostat and detector materials are reduced to negligible amounts by target fiducialization, hence we neglect them here [44]. The differential energy spectra of the above four ER background contributions are shown in Fig. 5 (left panel), both before and after detector-level event cuts.

**NR backgrounds:** Radiogenic neutrons emitted from the detector's materials, particularly from light PTFE used as insulator and light reflector, and photosensors made from various materials constitute a primary source of NR background.<sup>6</sup> Fiducialization of the detector volume serves as the primary detector-level cut on the radiogenic neutrons, which extensive `Geant4` simulations indicate as interacting primarily near the detector walls. Furthermore, neutrons can scatter multiple times within the detector volume. A veto on such multi-scatter events, determined from the S2 area distribution, is implemented with an assumed 100% efficiency.

The neutron background contributes more at larger (10–50 keV) recoil energies relative to the significantly more perilous other NR backgrounds, namely, coherent elastic neutrino-nucleus scattering (CEvNS) [44].  $^8\text{B}$  solar neutrinos are primarily responsible for a steep rise in background events at low recoil energy, hindering the detection of low-mass WIMPs (5–8 GeV). This background is difficult to distinguish from WIMP signals and represents a limit on sensitivity [62], at least for non-directional DD experiments.

At higher recoil energies, the main CEvNS background is from atmospheric neutrinos (atm), with smaller contributions from solar neutrinos from the helium-proton reaction (hep) and the diffuse supernova neutrino background (DSNB) [41, 63]. The spectra of NR backgrounds considered in this study are shown in Fig. 5 (right panel).

Lastly, as noted in Sect. 3.1, by propagating nuisance-parameter variations through the simulation and training on the resulting samples, the neural network effectively learns to marginalize over these uncertainties. For the generation of the PMT and waveform data used in this study, such nuisance parameters included within the `Tray` framework include PMT-quantum efficiencies, Light collection efficiency (LCE) uncertainties, systematic uncertainties associated with the S1/S2 reconstruction, as well as single photo-electron (SPE) area response.

<sup>6</sup> Work is currently being undertaken to improve the understanding of radiogenic neutrons in DARWIN as well as the uncertainty on their contribution. The resulting insights could very easily be included in the pipeline presented in this work in a future iteration.

## Appendix B Neural network architectures

In this section, we describe the components of the neural networks in detail. All neural networks are trained with `Tensorflow v2.15.0` [64].

### Appendix B.1 Variational autoencoder (VAE)

The goal of an autoencoder is to learn a compressed representation (encoding) of the input data, and then reconstruct the input data from this encoding [65, 66]. Autoencoders encompass three primary components: an encoder, a latent space, and a decoder. The encoder reduces the input data vectors  $\mathbf{x}_{\text{in}} \in \mathbb{R}^n$  into a lower-dimensional latent space representation  $\mathbf{z} \in \mathbb{R}^m$  (with  $m \ll n$ ) through a transformation  $\mathbf{z} = f(\mathbf{x})$ . The decoder then reconstructs the input from this compressed form, aiming to produce an output  $\mathbf{x}_D = g(\mathbf{z})$  as close to the original  $\mathbf{x}_{\text{in}}$  as possible. A reconstruction loss function, quantifying the difference between  $\mathbf{x}_{\text{in}}$  and  $\mathbf{x}_D$ , is optimized during training.

Variational Autoencoders (VAEs) extend this concept by introducing a probabilistic approach to the encoding process. Unlike standard autoencoders, the encoder in a VAE maps input data to a probability distribution characterized by mean  $\mu$  and variance  $\sigma^2$ , essentially transforming the encoder's output into the parameters of a Gaussian distribution:

$$f(\mathbf{x}_{\text{in}}) \rightarrow q(\mathbf{z} | \mathbf{x}_{\text{in}}) = \mathcal{N}_{\mathbf{z}}(\boldsymbol{\mu}, \text{diag}(\boldsymbol{\sigma}^2)).$$

The decoder, now governed by  $g(\mathbf{z}) \rightarrow p(\mathbf{x}_D | \mathbf{z})$ , is a probabilistic distribution that reconstructs data from sampled points in this probabilistic latent space. When  $\mathbf{x}_D$  are real vectors,  $p(\mathbf{x}_D | \mathbf{z})$  is taken to be a multidimensional normal distribution with diagonal covariant structure<sup>7</sup> [68]:

$$p(\mathbf{x}_D | \mathbf{z}) = \mathcal{N}_{\mathbf{x}_{\text{in}}}(\mathbf{x}_D, \text{diag}(\boldsymbol{\sigma}_D^2)). \quad (\text{B.1})$$

The VAE is trained via stochastic gradient descent by maximizing the loss function given by the so-called 'evidence lower bound' or ELBO [68]:

$$\begin{aligned} \text{ELBO} &= \mathbb{E}_{q(\mathbf{z} | \mathbf{x}_{\text{in}})}[\log p_{\mathbf{x}_{\text{in}}}(\mathbf{x}_D | \mathbf{z})] - D_{\text{KL}}(q(\mathbf{z} | \mathbf{x}_{\text{in}}) || p(\mathbf{z})) \\ &= \frac{1}{L} \sum_{l=1}^L \log \mathcal{N}_{\mathbf{x}_{\text{in},l}}(\mathbf{x}_l^D, \text{diag}(\boldsymbol{\sigma}_l^D)^2) \\ &\quad + \frac{1}{2} \beta \sum_{j=1}^m \left( 1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2 \right), \end{aligned} \quad (\text{B.2})$$

<sup>7</sup> This is a simplifying choice for the covariance structure. See Ref. [67] for an application of a structured Gaussian as the decoder.

**Table 2** Summary of the VAE architecture and optimal hyperparameters. All dense layers have ReLU activations except for the linear input and output.

Variational autoencoder architecture	
Latent dimension, $m$	128
$\beta$	10
Encoder	Input Layer: Shape: 3835 Dense Layer: 2000 units Dense Layer: 500 units Dense Layer: $2 \times m$
Decoder	Input Layer: Shape: $m$ Dense Layer: 500 units ( $\times 2$ ) Dense Layer: 2000 units ( $\times 2$ ) Dense Layer: 3835 units ( $\times 2$ )
Optimizer	Adamax, Learning Rate: 0.0005
Training epochs	30

where  $m$  is the dimensionality of the latent space (number of independent Gaussians), the expectation is under the distribution  $q(\mathbf{z} \mid \mathbf{x}_{\text{in}})$  and the data are batched into batches of size  $L$ . The coefficient  $\beta$  in the Kullback–Leibler (KL) term balances its regularization strength [69], with a higher  $\beta$  value ensuring that the encoded representations are closer to the prior,  $p(\mathbf{z})$ , taken to be a standard multivariate Gaussian,  $\mathcal{N}(\mathbf{0}_m, \mathbf{1}_m)$ .

The VAE architecture used in this study was selected after hyperparameter optimization on validation datasets withheld from training, and inspired by previously successful architectures in similar settings, in particular Ref. [32]. It consists of an encoder that takes vectorized data inputs  $\mathbf{x}_{\text{in}}$  (see Sect. 3.3) in batches of size  $L = 10$  and processes it through two dense (i.e., fully-connected) layers with 2000 and 500 units respectively. The latent space dimension is  $m = 128$ . The decoder has a dual-network structure. Both networks within the decoder begin with an input of shape 128, and process it through dense layers of 500 and 2000 units, culminating in two output layers  $\mathbf{x}_D$  and  $\log \sigma^2$  with shape matching  $\mathbf{x}_{\text{in}}$ . The architecture is summarized in Table 2. We note that this architecture does not scale for use on raw time series PMT readout data, given that dense, fully connected neural networks are not optimal for the sparsity one would expect from such data (leading to optimization issues and computational inefficiency). Therefore, a more suitable architecture, such as the one presented in Ref. [70], would be needed to use raw data as input.

For training, we use an Adamax optimizer with a learning rate of  $0.5 \times 10^{-3}$ . The entire training regimen is set to run for 30 epochs, with an optimized  $\beta$  value of 10 (via uniform hyperparameter scans). Validation tests of the so-obtained latent space representation are presented in Appendix E.

**Table 3** Summary of the neural network classifier's architecture and optimal hyperparameters. The activation function of all layers is ReLU, with the output being a sigmoid

Classifier Architecture	
Input Shape	Data Shape (3835)
Layers	Dense Layer: 256 units Dense Layer: 64 units Dense Layer: 16 units Output Layer: 1 unit
Optimizer	Adam, Learning Rate: 0.01
Training Epochs	5

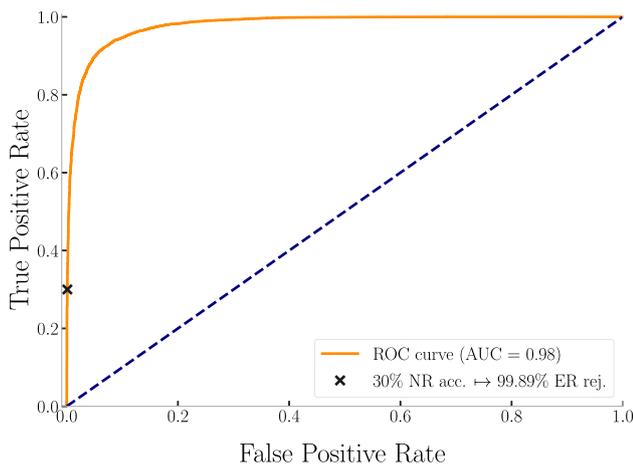
## Appendix B.2 Supervised ER vs NR classifier

The second component of the anomaly detector pipeline is a simple multi-layer perceptron (MLP) feed-forward neural network [71], whose architecture details are listed in Table 3. The classifier's task is to perform a binary classification between ER (0 output value) and NR (output value of 1) events. The MLP is trained by minimizing the standard binary cross-entropy loss:

$$H_B = -\frac{1}{L} \sum_{i=1}^L \log(1 - p(\mathbf{x}_{\text{in}})) \quad (\text{B.3})$$

where  $L = 10$  is the number of samples in the batch, and  $p(\mathbf{x}_{\text{in}})$  is the predicted class probability for each sample extracted from the sigmoid output of the MLP. The architecture details of this classifier are listed in Table 3.

In Fig. 6 we show the receiver operating characteristic curve (ROC) for a test set of  $10^4$  ER and NR events. We observe an area under the curve (AUC) of 0.98. We compare our classifier's performance with the 99.98% ER rejection obtained in previous DARWIN sensitivity studies [8, 9]. Such a large ER rejection probability is intended to mitigate leakage of ER's into the WIMP NR signal region, but it comes at the expense of a lower NR acceptance, which is estimated at 30% in Ref. [9]. The false positive rate (FPR) for our classifier (which corresponds to ER leakage) at a true positive rate (TPR) of 0.3 (correct NR classification), is denoted by the black cross in Fig. 6, and is 0.11%, corresponding to 99.89% ER rejection. We note that the standard approach uses the assumption of Gaussianity for cS1 and cS2 to mitigate ER leakage. Our classifier, however, makes no such assumption as it operates on an event-by-event basis, and hence any misclassified events will simply modify the distribution of the anomaly score, see Sec. 4. We also tried modifications of the loss function in Eq. (B.3) to optimize the false positive rate (i.e., minimize the number of mis-classified ER). Whilst this indeed was successful, the number of mis-classified NRs also



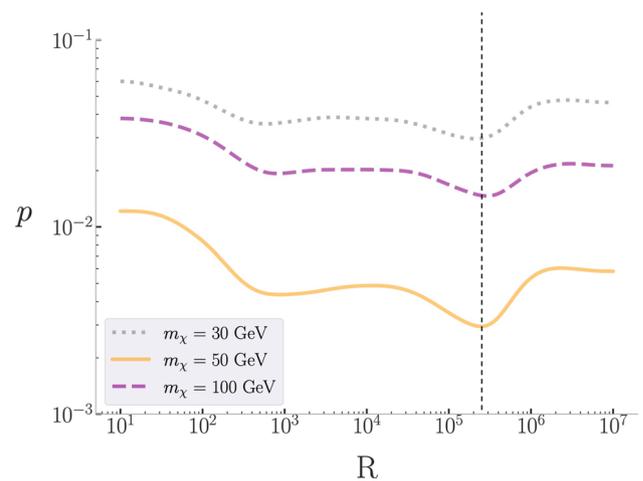
**Fig. 6** Receiver operating characteristic (ROC) curve of the supervised classifier trained to discriminate ER vs NR events, evaluated on a testing set consisting of an evenly mixed sample of  $10^4$  NR and  $10^4$  ER events. The area under the curve (AUC) is 0.98. The dashed blue lines indicate a random classifier. The black cross denotes the false positive rate (FPR) at a true positive rate (TPR) of 0.3, corresponding to the ER rejection capability of the classifier when the NR acceptance is 30%

increased, leading to a net zero effect in the overall anomaly score presented in Eqn. (2). Finally, we note that the choice of architecture used to perform the event-by-event ER/NR classification is largely irrelevant, as was seen in Ref. [36], where comparable performance was observed from MLPs, transformers, and boosted decision trees.

### Appendix C Optimization of the supervised contribution

The optimization of the hyperparameter  $R$  in Eq. (2) is a choice to be made at the time of analysis, in order to maximize the observation of any anomalous  $TS$  component, if it exists. To demonstrate this, we perform a scan over a range of logarithmically spaced  $R$  values  $R \in [1, 10^7]$  at fixed signal injection benchmarks corresponding to WIMP masses of 30, 50 and 100 GeV at an exposure of 200ty. These values of the WIMP mass were chosen in order to vary the spectral dependence of the induced WIMP signal. We show the median sensitivity, defined in Eq. (4), in Fig. 7, as a function of  $R$ , for the three benchmarks. A smaller  $p$ -value means better anomaly awareness and higher power to reject  $\mathcal{H}_0$  in the presence of a signal, and thus  $R$  should be chosen to minimize this value. We conduct this test at a cross-section that yields a background rejection  $p$ -value of at least  $\sim 3\sigma$  for  $m_\chi = 50$  GeV, so as to have ample statistics to perform the test for all three mass benchmarks.

We observe that the spectral dependence of the anomaly function  $TS$  entering through the ELBO as observed in Fig. 8 does not affect the dependence of the optimal  $R$  value, which

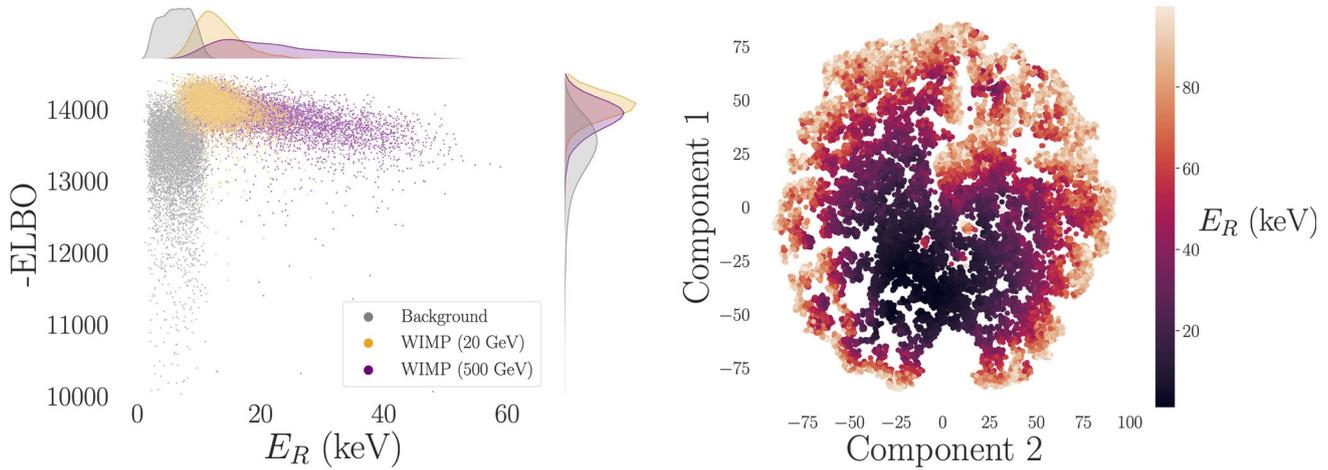


**Fig. 7** optimization of the hyperparameter  $R$  that controls the contribution of the supervised classifier in the determination of the anomaly score  $TS$ . The  $p$ -value to reject  $\mathcal{H}_0$  is given as a function of  $R$  for three benchmark WIMP masses at fixed exposure of 200 ty and cross-section  $\sigma_{SI} = 6.5 \times 10^{-48} \text{ cm}^2$ . As the scattering cross-section merely rescales the median sensitivity, the choice of  $R$  and cut value are insensitive to it. The optimal value for  $m_\chi = 50$  GeV is  $R = 2.5 \times 10^5$ , shown by the vertical dashed line. The variation in the location of optimal  $R$  value is small for other mass benchmarks

lies at  $\sim 2.5 \times 10^5$ . The general variability of the  $p$ -value is much more pronounced for  $m_\chi = 50$  GeV due to DARWINs elevated sensitivity to this mass. We observe that for  $R$  values above  $\sim 10^6$ , the  $p$ -value exhibits a plateau, which we have checked persists for values  $R > 10^7$ . This indicates that above this critical value of  $R$ , the influence of the VAE is vanishingly small.

The above results highlight the importance of taking a semi-supervised approach: the fact that the power to reject  $\mathcal{H}_0$  is maximized for  $R \neq 0, \infty$  shows explicitly the need for a combined supervised and unsupervised approach in order to maximize sensitivity to anomalous physics. In principle,  $R$  could be recast as a learnable parameter during training, although we chose to leave this to future study. For this work, we adopt an optimized  $R$  value of  $R = 2.5 \times 10^5$ .

Previous studies observed that classifiers can perform well as anomaly detectors (see, for example, Ref. [72]). An admixture of many supervised and/or unsupervised components could offer additional advantages, for example, by further exploiting the topological structure of events observed in the latent space. Indeed, Fig. 7 indicates non-triviality via the two observed local minima in the  $R$  dependence of the median sensitivity. We see that the latent data feature that is learned by the VAE was the event recoil energy, whilst the classifier learns the type of event. Both of these features are crucial to a new physics discovery, regardless of origin. It may then follow that other auxiliary models trained on the same and/or combinations/sets of prompt detector outputs may yield even



**Fig. 8** **Left:** ELBO values as a function of ground truth recoil energy  $E_R$  for a validation set of events: total (ER+NR) background (gray) and events from two WIMP benchmarks. The 1D marginals of the ELBO and  $E_R$  are also shown. Distributions are normalized densities for illustration purposes. The separation in the 2D space shows that spectral information has been encoded within the ELBO. **Right:** 2D tSNE of

the trained VAE’s 128-dimensional latent space for a validation sample of ER events with true recoil energies in the range  $E_R \in [1 - 100]$ . The color scale represents ground-truth recoil energy  $E_R$  of the events. The non-trivial latent structure in  $E_R$  confirms that the model has learned spectral information

better anomaly awareness. We leave this as an interesting question for future work in this domain.

**Appendix D Spectral information encoding**

The distribution of the ELBO from the VAE as a function of ground truth (simulated) event recoil energy  $E_R$  is shown for a validation sample with true recoil energies in the range [1–100] keV in Fig. 8 (left panel). We plot the normalized spectral distributions in the space spanned by  $E_R$  and ELBO for the total (ER+NR) background (gray) and for events generated by two WIMP masses,  $m_\chi = 20, 500$  GeV (orange/magenta). We observe an interesting separation in the distributions, pointing to the fact that the VAE is capable, after training exclusively on background events, to distinguish the spectral distribution from WIMPs.

To visualize the latent representation of the data, we further show a 2-dimensional t-distributed stochastic neighbor embedding (tSNE) projection [73] of the 128-dimensional latent space of the VAE in Fig. 8 (right panel). The non-trivial structure of the latent space, even in a two-dimensional projection, demonstrates that spectral information has indeed been incorporated into the model.

**Appendix E Validation of VAE representation**

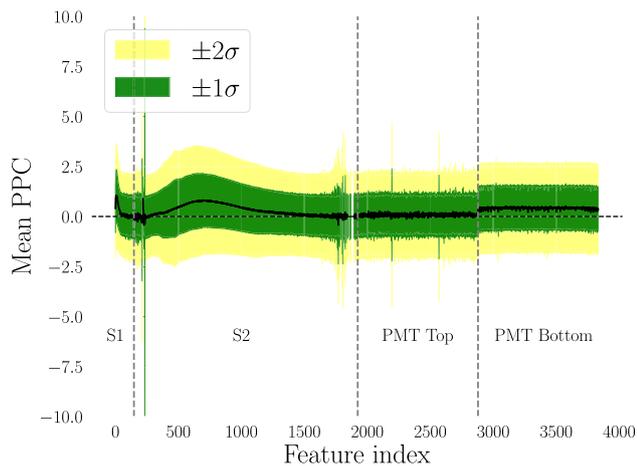
To validate the true low-dimensional latent features of the ER training data that were learned by the VAE, we carry out a standard benchmarking test known as a ‘posterior predictive check’ (PPC) [74, 75]. An ideal model will generatively

produce samples that align with the target distribution, and therefore produce a PPC  $\sim 0$ . Given the one-dimensional nature of our data (after vectorisation), we adopt the following simple strategy: we generate  $N$  samples  $\tilde{\mathbf{z}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$  from the latent space of the trained VAE and parsing them through the decoder, to obtain the predicted output,  $\tilde{\mathbf{x}}$ . A separate test set  $\mathbf{x}_{\text{test}}$  that is withheld from training is then used to calculate the relative reconstruction error:

$$\text{Mean (PPC)}^{(j)} = \frac{1}{N} \sum_{i=1}^N \frac{(\tilde{x}_i^{(j)} - x_{i_{\text{test}}}^{(j)})}{\sigma_{\text{test}}^{(j)}}, \tag{E.4}$$

where  $\sigma_{\text{test}}^{(j)} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{i_{\text{test}}}^{(j)} - \bar{x}_{\text{test}}^{(j)})^2}$  is the standard deviation of the distribution of test samples  $x_{\text{test}}^{(j)}$  for feature (vector column)  $j = 1, \dots, 3835$  and serves as a normalization factor. We use  $N = 10^4$ .

We show the result of the PPC in Fig. 9 for all 3835 data features. For clarity, we demarcate with vertical lines the features corresponding to the S1 and S2 waveforms, as well as the top and bottom S2 PMT hit-patterns. We plot the mean PPC as a black curve with the  $\pm 1, 2\sigma$  uncertainties in green and yellow, respectively. While a perfect network would produce a PPC of zero for all features, we observe that our network’s output lies within  $1\sigma$  of 0 for all features, indicating that it has learnt the underlying properties of the training data. The largest deviation from zero occurs for features at small pulse times for S1 and S2 waveforms (i.e., close to the start of the S1/S2 feature indices, indicated by the vertical lines). This is expected since most of the events used during training have a small or zero S1/S2 value at larger times (cf. Figure 1).



**Fig. 9** Posterior predictive checks performed on  $10^4$  samples from the latent space of the trained VAE. A perfect VAE would produce a PPC of zero for all feature indices. The black curve is the mean PPC from Eqn. (E.4), with  $\pm 1\sigma$  and  $2\sigma$  estimates shown as green and yellow bands, respectively. Each feature index corresponds to an element of the input data vector  $\mathbf{x}_{in}$ . The vertical gray dashed lines demarcate the subdivision into the S1/S2 wave-forms and S2 PMT Top and PMT Bottom hit patterns

Therefore, the network has fewer issues learning this degeneracy at large times and can reconstruct the corresponding features toward the ends of the S1/S2 feature index. This, however, can lead to larger variance in the PPC distribution of some features due to the model's lack of reconstruction power in regions of degenerate zeros in the feature space, as are observed as spikes in the  $1/2\sigma$  bands. We observe near-perfect reconstruction for the top S2 PMT array, but observe a slight, positive offset for the bottom PMT. We attribute this behavior to the bottom PMT displaying what is mostly uniform noise for the majority of ER events, as seen in Fig. 1. Hence, the values for which the VAE can optimize the ELBO are somewhat arbitrary and present as a systematic offset. The top PMT array, however, displays concentrated deposits that are well associated with the event properties and can therefore be learned more easily.

## References

1. M.W. Goodman, E. Witten, *Phys. Rev. D* **31**, 3059 (1985). <https://doi.org/10.1103/PhysRevD.31.3059>
2. E. Aprile et al., *Phys. Rev. Lett.* **131**(4), 041003 (2023). <https://doi.org/10.1103/PhysRevLett.131.041003>
3. LUX-ZEPLIN-collaboration, *Physical Review Letters* **131**(4) (2023). <https://doi.org/10.1103/physrevlett.131.041002>
4. Y. Meng et al., *Phys. Rev. Lett.* **127**(26), 261802 (2021). <https://doi.org/10.1103/PhysRevLett.127.261802>
5. M. Lai, *JINST* **18**(02), C02046 (2023). <https://doi.org/10.1088/1748-0221/18/02/C02046>
6. C. Aalseth et al., *Eur. Phys. J. Plus* **133**, 131 (2018). <https://doi.org/10.1140/epjp/i2018-11973-4>

7. J. Calvo et al., *JCAP* **03**, 003 (2017). <https://doi.org/10.1088/1475-7516/2017/03/003>
8. J. Aalbers et al., *J. Cosmol. Astropart. Phys.* **2016**(11), 017–017 (2016). <https://doi.org/10.1088/1475-7516/2016/11/017>
9. J. Aalbers et al., *J. Phys. G: Nucl. Part. Phys.* **50**(1), 013001 (2022). <https://doi.org/10.1088/1361-6471/ac841a>
10. J. Aalbers, et al., (2024). <https://inspirehep.net/literature/2841888>
11. G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, L. Zdeborová, *Rev. Mod. Phys.* **91**(4), 045002 (2019). <https://doi.org/10.1103/RevModPhys.91.045002>
12. X. Zhang, Y. Wang, W. Zhang, Y. Sun, S. He, G. Contardo, F. Villaescusa-Navarro, S. Ho, (2019). <https://ui.adsabs.harvard.edu/abs/2019arXiv190205965Z/abstract>
13. L. Lucie-Smith, H.V. Peiris, A. Pontzen, *Mon. Not. Roy. Astron. Soc.* **490**(1), 331 (2019). <https://doi.org/10.1093/mnras/stz2599>
14. M. Bernardini, L. Mayer, D. Reed, R. Feldmann, *Mon. Not. Roy. Astron. Soc.* **496**(4), 5116 (2020). <https://doi.org/10.1093/mnras/staa1911>
15. E. Todarello, A. Scaffidi, M. Regis, M. Taoso, *JCAP* **01**, 062 (2024). <https://doi.org/10.1088/1475-7516/2024/01/062>
16. M. Farina, Y. Nakai, D. Shih, *Phys. Rev. D* **101**, 075021 (2020). <https://doi.org/10.1103/PhysRevD.101.075021>
17. L.M. Dery, B. Nachman, F. Rubbo, A. Schwartzman, *J. Phys: Conf. Ser.* **1085**(4), 042006 (2018). <https://doi.org/10.1088/1742-6596/1085/4/042006>
18. J.H. Collins, K. Howe, B. Nachman, *Phys. Rev. Lett.* **121**(24), 241803 (2018). <https://doi.org/10.1103/PhysRevLett.121.241803>
19. S. Otten, S. Caron, W. de Swart, M. van Beekveld, L. Hendriks, C. van Leeuwen, D. Podareanu, R. Ruiz de Austri, R. Verheyen, *Nature Commun.* **12**(1), 2985 (2021). <https://doi.org/10.1038/s41467-021-22616-z>
20. A. Blance, M. Spannowsky, P. Waite, *Journal of High Energy Physics* **2019**(10) (2019). [https://doi.org/10.1007/jhep10\(2019\)047](https://doi.org/10.1007/jhep10(2019)047)
21. A. Blance, M. Spannowsky, *Journal of High Energy Physics* **2021**(2) (2021). [https://doi.org/10.1007/jhep02\(2021\)212](https://doi.org/10.1007/jhep02(2021)212)
22. T. HeimeI, G. Kasieczka, T. Plehn, J.M. Thompson, *Sci. Post. Phys* **6**(3), 030 <https://doi.org/10.21468/SciPostPhys.6.3.030>
23. M. Kuusela, T. Vatanen, E. Malmi, T. Raiko, T. Aaltonen, Y. Nagai, *J. Phys: Conf. Ser.* **368**, 012032 (2012). <https://doi.org/10.1088/1742-6596/368/1/012032>
24. O. Knapp, G. Dissertori, O. Cerri, T.Q. Nguyen, J.R. Vlimant, M. Pierini, arXiv preprint [arXiv:2005.01598](https://arxiv.org/abs/2005.01598) (2020). <https://link.springer.com/article/10.1140/epjp/s13360-021-01109-4>
25. A. Andreassen, B. Nachman, D. Shih, *Phys. Rev. D* **101**(9), 095004 (2020). <https://doi.org/10.1103/PhysRevD.101.095004>
26. B. Nachman, D. Shih, *Phys. Rev. D* **101**, 075042 (2020). <https://doi.org/10.1103/PhysRevD.101.075042>
27. J.H. Collins, K. Howe, B. Nachman, *Phys. Rev. D* **99**(1), 014038 (2019). <https://doi.org/10.1103/PhysRevD.99.014038>
28. O. Cerri, T.Q. Nguyen, M. Pierini, M. Spiropulu, J.R. Vlimant, *J. High Energy Phys.* **2019**(5), 36 (2019). [https://doi.org/10.1007/JHEP05\(2019\)036](https://doi.org/10.1007/JHEP05(2019)036)
29. M. van Beekveld, S. Caron, L. Hendriks, P. Jackson, A. Leinweber, S. Otten, R. Patrick, R. Ruiz De Austri, M. Santoni, M. White, Combining outlier analysis algorithms to identify new physics at the LHC (2021). [https://doi.org/10.1007/JHEP09\(2021\)024](https://doi.org/10.1007/JHEP09(2021)024)
30. C.K. Khosa, V. Sanz, *SciPost. Phys* **15**, 053 (2023). <https://doi.org/10.21468/SciPostPhys.15.2.053>
31. I. Coarasa et al., *JCAP* **11**, 048 (2022). <https://doi.org/10.1088/1475-7516/2022/11/048>. ([Erratum: *JCAP* **06**, E01 (2023)])
32. J. Herrero-Garcia, R. Patrick, A. Scaffidi, *J. Cosmol. Astropart. Phys.* **2022**(02), 039 (2022). <https://doi.org/10.1088/1475-7516/2022/02/039>

33. D.S. Akerib, others (LUX Collaboration), *Physical Review D* **106**(7), 072009 (2022). <https://doi.org/10.1103/PhysRevD.106.072009>
34. P. Agnes et al., *Eur. Phys. J. C* **83**, 322 (2023). <https://doi.org/10.1140/epjc/s10052-023-11410-4>
35. E. Aprile et al., *Phys. Rev. D* **108**(1), 012016 (2023). <https://doi.org/10.1103/PhysRevD.108.012016>
36. D.E. López-Fogliani, A.D. Perez, R.R. de Austri, *J. Cosmol. Astropart. Phys.* **2025**(01), 057 (2025). <https://doi.org/10.1088/1475-7516/2025/01/057>
37. V. Chepel, H. Araújo, *J. Instrum* **8**(04), R04001 (2013). <https://doi.org/10.1088/1748-0221/8/04/R04001/pdf>
38. R. Peres, *J. Instrum.* **18**(03), C03027 (2023). <https://doi.org/10.1088/1748-0221/18/03/c03027>
39. A. Brown, H. Fischer, R. Glade-Beucke, J. Grigat, F. Kuger, S. Lindemann, T. Luce, D. Masson, J. Müller, J. Reininghaus, M. Schumann, A. Stevens, F. Tönnies, F. Toschi, *J. Instrum.* **19**(05), P05018 (2024). <https://doi.org/10.1088/1748-0221/19/05/P05018>
40. S. Agostinelli et al., *Nucl. Instrum. Meth. A* **506**, 250 (2003)
41. DARWIN-Collaboration, *The European Physical Journal C* **84**(1) (2024). <https://doi.org/10.1140/epjc/s10052-023-12298-w>
42. Nest version v2.3.12 (2018). <https://doi.org/10.5281/zenodo.1314499>
43. I. collaboration, *Journal of Instrumentation* **17**(06), P06026 (2022). <https://doi.org/10.1088/1748-0221/17/06/p06026>
44. M. Schumann, L. Baudis, L. Bütikofer, A. Kish, M. Selvi, *J. Cosmol. Astropart. Phys.* **2015**(10), 016 (2015). <https://doi.org/10.1088/1475-7516/2015/10/016>
45. M. Weber, Gentle neutron signals and noble background in the xenon100 dark matter search experiment. Ph.D. thesis, Ruprecht-Karls-Universität Heidelberg (2013). <https://core.ac.uk/download/pdf/161443046.pdf>
46. G. Kessler, in *20th International Conference on Particles and Nuclei* (2014), pp. 357–360. <https://doi.org/10.3204/DESY-PROC-2014-04/109>
47. E. Aprile et al., *JINST* **18**(07), P07054 (2023). <https://doi.org/10.1088/1748-0221/18/07/P07054>
48. C. Hewitt, M. Anderson, in *Neutrino Physics and Machine Learning 2024* (ETH Zurich, 2024). <https://indico.phys.ethz.ch/event/113/contributions/827/>
49. W. Jiang, G. Huang, Z. Liu, W. Luo, L. Wen, J. Luo, (2024). <https://link.springer.com/article/10.1140/epjc/s10052-024-13724-3>
50. S. Farrell, M. Bergevin, A. Bernstein, (2024). <https://ui.adsabs.harvard.edu/abs/2024arXiv240706139F/abstract>
51. U. Zivanovic, S. Di Gioia, A. Scaffidi, M. de los Rios, G. Contardo, R. Trotta, arXiv e-prints [arXiv:2505.20535](https://arxiv.org/abs/2505.20535) (2025).
52. S. Vetter, in *Neutrino Physics and Machine Learning (NPML)* (ETH Zurich, 2024). <https://indico.phys.ethz.ch/event/113/contributions/890/>
53. M. Arthurs, in *Conference on Science at the Sanford Underground Research Facility* (SD Mines, South Dakota, USA, 2024). <https://indico.sanfordlab.org/event/68/contributions/1323/>
54. R.L. Workman, et al., *PTEP* **2022**, 083C01 (2022). <https://doi.org/10.1093/ptep/ptac097>. <https://academic.oup.com/ptep/article/2022/8/083C01/6651666>
55. G. Cowan, K. Cranmer, E. Gross, O. Vitells, *Eur. Phys. J. C* **71**, 1554 (2011)
56. J. Hermans, A. Delaunoy, F. Rozet, A. Wehenkel, V. Begy, G. Louppe. A trust crisis in simulation-based inference? your posterior approximations can be unfaithful (2022). [arXiv:https://arxiv.org/abs/2110.06581](https://arxiv.org/abs/2110.06581)
57. D.E. and, *Journal of Statistical Computation and Simulation* **22**(3-4), 307 (1985). <https://doi.org/10.1080/00949658508810853>
58. D.S. Akerib, et al., (2016). [https://www.researchgate.net/publication/306285462\\_Low-energy\\_07-74\\_keV\\_nuclear\\_recoil\\_calibration\\_of\\_the\\_LUX\\_dark\\_matter\\_experiment\\_using\\_D-D\\_neutron\\_scattering\\_kinematics](https://www.researchgate.net/publication/306285462_Low-energy_07-74_keV_nuclear_recoil_calibration_of_the_LUX_dark_matter_experiment_using_D-D_neutron_scattering_kinematics)
59. B. Lenardo, et al., (2019). [https://www.researchgate.net/publication/301572459\\_Improved\\_Limits\\_on\\_Scattering\\_of\\_Weakly\\_Interacting\\_Massive\\_Particles\\_from\\_Reanalysis\\_of\\_2013\\_LUX\\_Data](https://www.researchgate.net/publication/301572459_Improved_Limits_on_Scattering_of_Weakly_Interacting_Massive_Particles_from_Reanalysis_of_2013_LUX_Data)
60. M. Gong, K. Zhang, B. Huang, C. Glymour, D. Tao, K. Batmanghelich, *ArXiv abs/1804.04333* (2018). <https://api.semanticscholar.org/CorpusID:4807438>
61. E. Aprile, et al., (2024). <https://doi.org/10.48550/arXiv.2406.13638>. <https://hal.science/hal-04659687>
62. C.A.J. O'Hare, *Phys. Rev. D* **94**(6), 063527 (2016). <https://doi.org/10.1103/PhysRevD.94.063527>
63. L.E. Strigari, *New J. Phys.* **11**(10), 105011 (2009). <https://doi.org/10.1088/1367-2630/11/10/105011>
64. M. Abadi, et al. (2015). <https://www.tensorflow.org/>
65. D. Bank, N. Koenigstein, R. Giryes. Autoencoders (2021). [https://doi.org/10.1007/978-3-031-24628-9\\_16](https://doi.org/10.1007/978-3-031-24628-9_16)
66. J. Schmidhuber, *Neural Netw.* **61**, 85 (2015). <https://doi.org/10.1016/j.neunet.2014.09.003>
67. G. Dorta, S. Vicente, L. de Agapito, N.D.F. Campbell, I.J.A. Simpson, (2018). <https://api.semanticscholar.org/CorpusID:4560603>
68. D.P. Kingma, M. Welling. Auto-encoding variational bayes (2022). <http://web2.cs.columbia.edu/~blei/fogm/2018F/materials/KingmaWelling2013.pdf>
69. H. Sikka, W. Zhong, J. Yin, C. Pehlevant, in *2019 53rd Asilomar Conference on Signals, Systems, and Computers* (2019), pp. 888–895. <https://doi.org/10.1109/IEEECONF44664.2019.9048921>. [https://ieeexplore.ieee.org/abstract/document/9048921?casa\\_token=beCDJ95S9SsAAAAA:yG8VCLJ40\\_QZ\\_Oa7XVQpW3Te4G2xHBKiUKskBYqqQNIPf1PlrvoQmKYEaPBL3DcakoyGeAFW9Q](https://ieeexplore.ieee.org/abstract/document/9048921?casa_token=beCDJ95S9SsAAAAA:yG8VCLJ40_QZ_Oa7XVQpW3Te4G2xHBKiUKskBYqqQNIPf1PlrvoQmKYEaPBL3DcakoyGeAFW9Q)
70. A. Kumar, A. Singh, K. Doshi, A. Goyal, arXiv preprint [arXiv:2209.14249](https://arxiv.org/abs/2209.14249) (2022)
71. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016). <http://www.deeplearningbook.org>
72. L. Builtjes, S. Caron, P. Moskvitina, C. Nellist, R.R. de Austri, R. Verheyen, Z. Zhang. Attention to the strengths of physical interactions: Transformer and graph-based event classification for particle physics experiments (2024). <https://inspirehep.net/literature/2180384>
73. L. van der Maaten, G. Hinton, *J. Mach. Learn. Res.* **9**(86), 2579 (2008)
74. D. Mimno, D.M. Blei, B.E. Engelhardt, *Proc. Natl. Acad. Sci.* **112**(26), E3441 (2015). <https://doi.org/10.1073/pnas.1412301112>
75. A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, (2013). <https://sites.stat.columbia.edu/gelman/book/BDA3.pdf>

**XLZD Collaboration\***

J. Aalbers<sup>1</sup> , K. Abe<sup>2</sup> , M. Adrover<sup>3</sup>, S. Ahmed Maouloud<sup>4</sup> , L. Althueser<sup>5</sup> , D. W. P. Amaral<sup>6</sup> , B. Andrieu<sup>4</sup> , E. Angelino<sup>7,8</sup> , D. Antón Martín<sup>9</sup>, B. Antunovic<sup>10,a</sup>, E. Aprile<sup>11</sup> , M. Babicz<sup>3</sup>, D. Bajpai<sup>12</sup>, M. Balzer<sup>13</sup>, E. Barberio<sup>14</sup>, L. Baudis<sup>3</sup> , M. Bazyk<sup>14,15</sup> , N. F. Bell<sup>14</sup>, L. Bellagamba<sup>16</sup> , R. Biondi<sup>17</sup> , Y. Biondi<sup>18</sup>, A. Bismark<sup>3</sup> , C. Boehm<sup>19</sup>, K. Boese<sup>17</sup> , R. Braun<sup>5</sup>, A. Breskin<sup>20</sup>, S. Brommer<sup>21</sup>, A. Brown<sup>22,e</sup> , G. Bruni<sup>16</sup> , R. Budnik<sup>20</sup> , C. Cai<sup>25</sup>, C. Capelli<sup>3</sup> , A. Chauvin<sup>26</sup>, A. P. Cimental Chavez<sup>3</sup> , A. P. Colijn<sup>27</sup> , J. Conrad<sup>28</sup> , J. J. Cuenca-García<sup>3</sup> , V. D'Andrea<sup>8,b</sup> , L. C. Daniel Garcia<sup>4</sup> , M. P. Decowski<sup>27</sup> , A. Deisting<sup>29</sup> , C. Di Donato<sup>30</sup> , P. Di Gangi<sup>16</sup> , S. Diglio<sup>15</sup> , M. Doerenkamp<sup>26</sup>, G. Drexlin<sup>21</sup>, K. Eitel<sup>18</sup> , A. Elykov<sup>18</sup> , R. Engel<sup>18</sup>, A. D. Ferella<sup>8,30</sup> , C. Ferrari<sup>8</sup> , H. Fischer<sup>22</sup> , T. Flehmke<sup>28</sup> , M. Flierman<sup>27</sup> , K. Fujikawa<sup>31</sup>, W. Fulgione<sup>7,8</sup> , C. Fuselli<sup>27</sup> , P. Gaemers<sup>27</sup> , R. Gaior<sup>4</sup> , M. Galloway<sup>3</sup> , F. Gao<sup>25</sup> , N. Garroum<sup>4</sup>, R. Giacomobono<sup>32</sup> , F. Girard<sup>4</sup>, R. Glade-Beucke<sup>22</sup> , F. Glück<sup>18</sup>, L. Grandi<sup>9</sup> , J. Grigat<sup>22</sup> , R. Größle<sup>18</sup>, H. Guan<sup>33</sup> , M. Guida<sup>17</sup> , P. Gyorgy<sup>29</sup> , R. Hamann<sup>17</sup> , V. Hannen<sup>5</sup>, S. Hansmann-Menzemer<sup>26</sup>, N. Hargittai<sup>20</sup>, A. Higuera<sup>6</sup> , C. Hils<sup>29</sup>, K. Hiraoka<sup>31</sup>, L. Hoetzsch<sup>17</sup> , N. F. Hood<sup>35</sup> , M. Iacovacci<sup>32</sup> , Y. Itow<sup>31</sup> , J. Jakob<sup>5</sup>, R. S. James<sup>14,36</sup>, F. Joerg<sup>17,3</sup> , F. Kahlert<sup>33</sup> , Y. Kaminaga<sup>2</sup> , M. Kara<sup>18</sup> , P. Kavargin<sup>20</sup> , S. Kazama<sup>31</sup> , M. Keller<sup>26</sup>, P. Kharbanda<sup>27</sup>, B. Kilminster<sup>3</sup>, M. Kleifges<sup>13</sup>, M. Klute<sup>21</sup>, M. Kobayashi<sup>31</sup>, D. Koke<sup>5</sup>, A. Kopec<sup>37</sup> , B. von Krosigk<sup>38</sup>, F. Kuger<sup>22</sup> , L. LaCascio<sup>21</sup>, H. Landsman<sup>20</sup> , R. F. Lang<sup>33</sup> , L. Levinson<sup>20</sup> , I. Li<sup>6</sup> , A. Li<sup>35</sup>, S. Li<sup>39</sup> , S. Liang<sup>6</sup> , Z. Liang<sup>40</sup>, Y. -T. Lin<sup>17</sup>, S. Lindemann<sup>22</sup> , M. Lindner<sup>17</sup> , K. Liu<sup>25</sup>, J. Loizeau<sup>15</sup>, F. Lombardi<sup>29</sup> , J. Long<sup>9</sup> , J. A. M. Lopes<sup>41,c</sup> , G. M. Lucchetti<sup>16</sup> , T. Luce<sup>22</sup>, Y. Ma<sup>35</sup> , C. Macolino<sup>8,30</sup> , J. Mahlstedt<sup>28</sup> , B. Maier<sup>21,42</sup>, A. Mancuso<sup>16</sup> , L. Manenti<sup>19</sup> , F. Marignetti<sup>32</sup> , K. Martens<sup>2</sup> , J. Masbou<sup>15</sup> , E. Masson<sup>4</sup> , S. Mastroianni<sup>32</sup> , A. Melchiorre<sup>30</sup> , J. Menéndez<sup>43</sup> , M. Messina<sup>8</sup> , B. Milosovic<sup>10</sup>, S. Milutinovic<sup>10</sup>, K. Miuchi<sup>44</sup> , R. Miyata<sup>31</sup>, A. Molinaro<sup>7</sup> , C. M. B. Monteiro<sup>14</sup>, K. Morà<sup>11</sup> , S. Moriyama<sup>2</sup> , E. Morteau<sup>15</sup>, Y. Mosbacher<sup>20</sup>, J. Müller<sup>22</sup> , M. Murra<sup>11</sup> , J. L. Newstead<sup>14</sup>, K. Ni<sup>35</sup> , C. O'Hare<sup>19</sup>, U. Oberlack<sup>29</sup> , M. Obradovic<sup>10</sup>, I. Ostrowskiy<sup>12</sup>, S. Ouahada<sup>3</sup>, B. Paetsch<sup>20</sup> , Y. Pan<sup>4</sup> , M. Pandurovic<sup>10</sup>, Q. Pellegrini<sup>4</sup> , R. Peres<sup>3</sup> , F. Piastra<sup>3</sup> , J. Pienaar<sup>9,20</sup> , M. Pierre<sup>27</sup> , G. Plante<sup>11</sup> , T. R. Pollmann<sup>27</sup> , L. Principe<sup>14,15</sup> , J. Qi<sup>35</sup> , K. Qiao<sup>27</sup>, J. Qin<sup>6</sup> , M. Rajado<sup>3</sup> , D. Ramírez García<sup>3</sup> , A. Ravindran<sup>14,15</sup> , A. Razeto<sup>8</sup>, L. Sanchez<sup>6</sup>, P. Sanchez-Lucas<sup>3,d</sup>, G. Sartorelli<sup>16</sup> , A. Scaffidi<sup>45</sup>, J. Schreiner<sup>17</sup>, P. Schulte<sup>5</sup> , H. Schulze EiBing<sup>5</sup> , M. Schumann<sup>22</sup> , A. Schwenck<sup>18</sup>, L. Scotto Lavina<sup>4</sup> , M. Selvi<sup>16</sup> , F. Semeria<sup>16</sup> , P. Shagin<sup>29</sup>, S. Sharma<sup>26</sup>, W. Shen<sup>26</sup>, S. Y. Shi<sup>11</sup> , T. Shimada<sup>31</sup>, H. Simgen<sup>17</sup> , R. Singh<sup>33</sup> , M. Solmaz<sup>21,38</sup>, O. Stanley<sup>14,15</sup>, M. Steidl<sup>18</sup>, A. Stevens<sup>22</sup> , A. Takeda<sup>2</sup> , P.-L. Tan<sup>28</sup> , D. Thers<sup>15</sup> , T. Thümmler<sup>18</sup>, F. Tönnies<sup>22</sup> , F. Toschi<sup>18</sup>, G. Trinchero<sup>7</sup>, R. Trotta<sup>42,45</sup> , C. D. Tunnell<sup>6</sup> , P. Urquijo<sup>14</sup>, M. Utoyama<sup>31</sup>, K. Valerius<sup>18</sup> , S. Vecchi<sup>47</sup> , S. Vetter<sup>18</sup> , G. Volta<sup>17</sup> , D. Vorkapic<sup>10</sup>, W. Wang<sup>12</sup>, K. M. Weerman<sup>27</sup>, C. Weinheimer<sup>5</sup> , M. Weiss<sup>20</sup> , D. Wenz<sup>5</sup>, M. Wilson<sup>18</sup>, C. Wittweg<sup>3</sup> , J. Wolf<sup>21</sup>, V. H. S. Wu<sup>18</sup> , S. Wüstling<sup>13</sup>, M. Wurm<sup>29</sup>, Y. Xing<sup>14</sup> , D. Xu<sup>11</sup> , Z. Xu<sup>11</sup> , M. Yamashita<sup>2</sup> , L. Yang<sup>35</sup> , J. Ye<sup>40</sup> , L. Yuan<sup>9</sup> , G. Zavattini<sup>47</sup> , M. Zhong<sup>35</sup>, K. Zuber<sup>48</sup> 

<sup>1</sup> Nikhef and the University of Groningen, Van Swinderen Institute, Groningen 9747AG, The Netherlands

<sup>2</sup> Kamioka Observatory, Institute for Cosmic Ray Research, and Kavli Institute for the Physics and Mathematics of the Universe (WPI), University of Tokyo, Hida, Higashi-Mozumi, Kamioka, Gifu 506-1205, Japan

<sup>3</sup> Physik-Institut, University of Zürich, Zürich 8057, Switzerland

<sup>4</sup> LPNHE, Sorbonne Université, CNRS/IN2P3, Paris 75005, France

<sup>5</sup> Institute for Nuclear Physics, University of Münster, Münster 48149, Germany

<sup>6</sup> Department of Physics and Astronomy, Rice University, Houston, TX 77005, USA

<sup>7</sup> INAF-Astrophysical Observatory of Torino, Department of Physics, University of Torino and INFN-Torino, Turin 10125, Italy

<sup>8</sup> INFN-Laboratori Nazionali del Gran Sasso and Gran Sasso Science Institute, L'Aquila 67100, Italy

<sup>9</sup> Department of Physics, Enrico Fermi Institute & Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

<sup>10</sup> Vinca Institute of Nuclear Science, University of Belgrade, Mihajla Petrovica Alasa 12-14, Belgrade, Serbia

<sup>11</sup> Physics Department, Columbia University, New York, NY 10027, USA

<sup>12</sup> Department of Physics & Astronomy, University of Alabama, Tuscaloosa, AL 34587-0324, USA

<sup>13</sup> Institute for Data Processing and Electronics, Karlsruhe Institute of Technology, Karlsruhe 76021, Germany

<sup>14</sup> ARC Centre of Excellence for Dark Matter Particle Physics, School of Physics, The University of Melbourne, Vic 3010, Australia

<sup>15</sup> SUBATECH, IMT Atlantique, CNRS/IN2P3, Nantes Université, Nantes 44307, France

- <sup>16</sup> Department of Physics and Astronomy, University of Bologna and INFN-Bologna, Bologna 40126, Italy
- <sup>17</sup> Max-Planck-Institut für Kernphysik, Heidelberg 69117, Germany
- <sup>18</sup> Institute for Astroparticle Physics, Karlsruhe Institute of Technology, Karlsruhe 76021, Germany
- <sup>19</sup> School of Physics, The University of Sydney, Camperdown, Sydney, NSW 2006, Australia
- <sup>20</sup> Department of Particle Physics and Astrophysics, Weizmann Institute of Science, Rehovot 7610001, Israel
- <sup>21</sup> Institute of Experimental Particle Physics, Karlsruhe Institute of Technology, Karlsruhe 76021, Germany
- <sup>22</sup> Physikalisches Institut, Universität Freiburg, Freiburg 79104, Germany
- <sup>23</sup> Physikalisches Institut, Universität Freiburg, Freiburg, (Now at Sheffield) 79104, Germany
- <sup>24</sup> Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK
- <sup>25</sup> Department of Physics & Center for High Energy Physics, Tsinghua University, Beijing 100084, People's Republic of China
- <sup>26</sup> Physikalisches Institut, Universität Heidelberg, Heidelberg, Germany
- <sup>27</sup> Nikhef and the University of Amsterdam, Science Park, Amsterdam 1098XG, The Netherlands
- <sup>28</sup> Oskar Klein Centre, Department of Physics, Stockholm University, AlbaNova, Stockholm 10691, Sweden
- <sup>29</sup> Institut für Physik & Exzellenzcluster PRISMA+, Johannes Gutenberg-Universität Mainz, Mainz 55099, Germany
- <sup>30</sup> Department of Physics and Chemistry, University of L'Aquila, L'Aquila 67100, Italy
- <sup>31</sup> Kobayashi-Maskawa Institute for the Origin of Particles and the Universe, and Institute for Space-Earth Environmental Research, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan
- <sup>32</sup> Department of Physics "Ettore Pancini", University of Napoli and INFN-Napoli, Napoli 80126, Italy
- <sup>33</sup> Department of Physics and Astronomy, Purdue University, West Lafayette, IN 47907, USA
- <sup>34</sup> Albert Einstein Center for Fundamental Physics, Institute for Theoretical Physics, University of Bern, Sidlerstrasse 5, Bern 3012, Switzerland
- <sup>35</sup> Department of Physics, University of California San Diego, La Jolla, CA 92093, USA
- <sup>36</sup> Department of Physics and Astronomy, University College London (UCL), London WC1E 6BT, UK
- <sup>37</sup> Department of Physics & Astronomy, Bucknell University, Lewisburg, PA, USA
- <sup>38</sup> Kirchhoff-Institut für Physik, Universität Heidelberg, Heidelberg, Germany
- <sup>39</sup> Department of Physics, School of Science, Westlake University, Hangzhou 310030, People's Republic of China
- <sup>40</sup> School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Guangdong 518172, People's Republic of China
- <sup>41</sup> LIBPhys, Department of Physics, University of Coimbra, Coimbra 3004-516, Portugal
- <sup>42</sup> Physics Department, Imperial College London Blackett Laboratory, London SW7 2AZ, UK
- <sup>43</sup> Department of Quantum Physics and Astrophysics and Institute of Cosmos Sciences, University of Barcelona, Barcelona 08028, Spain
- <sup>44</sup> Department of Physics, Kobe University, Kobe, Hyogo 657-8501, Japan
- <sup>45</sup> Theoretical and Scientific Data Science, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste 34136, Italy
- <sup>46</sup> Department of Physics, Technische Universität Darmstadt, Darmstadt 64289, Germany
- <sup>47</sup> INFN-Ferrara and Dip. di Fisica e Scienze della Terra, Università di Ferrara, Ferrara 44122, Italy
- <sup>48</sup> Technische Universität Dresden, Dresden 01069, Germany
- <sup>a</sup> also at: University of Banja Luka, Banja Luka 78000, Bosnia and Herzegovina
- <sup>b</sup> also at: INFN-Roma Tre, Roma 00146, Italy
- <sup>c</sup> also at: Coimbra Polytechnic - ISEC, Coimbra 3030-199, Portugal
- <sup>d</sup> also at: University of Grenada, Grenada, Spain
- <sup>e</sup> Now at Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK