PROCEEDINGS OF SCIENCE

Trigger Optimization and Event Classification for Dark Matter Searches in the CYGNO Experiment Using Machine Learning

F D Amaro,[a] R Antonietti,[b,c] E Baracchini,[d,e] L Benussi,[f] C Capoccia,[f] M Caponero,[f,g] L G M de Carvalho,[h] G Cavoto,[i,j] I A Costa,[f] A Croce,[f] M D'Astolfo,[d,e] G D'Imperio,[j] G Dho,[f] E Di Marco,[j] J M F dos Santos,[a] D Fiorina,[d,e] F Iacoangeli,[j] Z Islam,[d,e] E Kemp,[k] H P Lima Jr,[d,e] G Maccarrone,[f] R D P Mano,[a] D J G Marques,[d,e] G Mazzitelli,[f] P Meloni,[b,c] A Messina,[i,j] C M B Monteiro,[a] R A Nobrega,[h] G M Oppedisano,[d,e,*] I F Pains,[h] E Paoletti,[f] F Petrucci,[b,c] S Piacentini,[d,e] D Pierluigi,[f] D Pinci,[j] F Renga,[j] A Russo,[f] G Saviano,[f,l] P A O C Silva,[a] N J Spooner,[m] R Tesauro,[f] S Tomassini[f] and D Tozzi[i,j]

[a] *LIBPhys, Department of Physics, University of Coimbra, 3004-516 Coimbra, Portugal*
[b] *Dipartimento di Matematica e Fisica, Università Roma Tre, 00146 Roma, Italy*
[c] *INFN Sezione di Roma Tre, 00146 Roma, Italy*
[d] *Gran Sasso Science Institute, 67100 L'Aquila, Italy*
[e] *INFN Laboratori Nazionali del Gran Sasso, 67100 Assergi, Italy*
[f] *INFN Laboratori Nazionali di Frascati, 00044 Frascati, Italy*
[g] *ENEA Centro Ricerche Frascati, 00044 Frascati, Italy*
[h] *Universidade Federal de Juiz de Fora, Faculdade de Engenharia, 36036-900 Juiz de Fora, MG, Brazil*
[i] *Dipartimento di Fisica, Sapienza Università di Roma, 00185 Roma, Italy*
[j] *INFN Sezione di Roma, 00185 Roma, Italy*
[k] *Universidade Estadual de Campinas (UNICAMP), Campinas 13083-859, SP, Brazil*
[l] *Dipartimento di Ingegneria Chimica, Materiali e Ambiente, Sapienza Università di Roma, 00185 Roma, Italy*
[m] *Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK*

E-mail: giuseppe.oppedisano@gssi.it

[*] Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).
All rights for text and data mining, AI training, and similar technologies for commercial purposes, are reserved.
ISSN 1824-8039. Published by SISSA Medialab.

PoS(14YRM2025)003

The CYGNO experiment employs an optical-readout Time Projection Chamber (TPC) to search for rare low-energy interactions using finely resolved scintillation images. While the optical readout provides rich topological information, it produces large, sparse megapixel images that challenge real-time triggering, data reduction, and background discrimination.

We summarize two complementary machine-learning approaches developed within CYGNO. First, we present a fast and fully unsupervised strategy for online data reduction based on reconstruction-based anomaly detection. A convolutional autoencoder trained exclusively on pedestal images (i.e. frames acquired with GEM amplification disabled) learns the detector noise morphology and highlights particle-induced structures through localized reconstruction residuals, from which compact Regions of Interest (ROIs) are extracted. On real prototype data, the selected configuration retains $(93.0 \pm 0.2)\%$ of reconstructed signal intensity while discarding $(97.8 \pm 0.1)\%$ of the image area, with ~ 25 ms per-frame inference time on a consumer GPU.

Second, we report a weakly supervised application of the Classification Without Labels (CWoLa) framework to data acquired with an Americium–Beryllium neutron source. Using only mixed AmBe and standard datasets (no event-level labels), a convolutional classifier learns to identify nuclear-recoil-like topologies. The achieved performance approaches the theoretical limit imposed by the mixture composition and isolates a high-score population with compact, approximately circular morphologies consistent with nuclear recoils.

14th Young Researcher Meeting (YRM 2025)
29 September – 3 October 2025
Gran Sasso Science Institute (GSSI), L'Aquila, Italy

PoS(14YRM2025)003

1. Introduction

Optical-readout Time Projection Chambers (TPCs) are powerful detectors for rare-event searches in the $O(1\text{–}100)$ keV regime, where short and localized nuclear-recoil (NR) tracks, the signature of dark-matter interactions, must be identified in the presence of abundant electronic-recoil (ER) backgrounds. In the CYGNO experiment [1], ionization electrons produced in a He–CF_4 gas mixture drift toward a triple Gas Electron Multiplier (GEM) stack, where they undergo charge amplification and induce CF_4 electroluminescence. This scintillation light is recorded by scientific CMOS cameras, while photomultiplier tubes provide timing information for three-dimensional reconstruction [2]. The optical readout offers excellent granularity and sensitivity to low-energy deposits [3].

A central challenge is data volume: each exposure is a megapixel-scale image in which the physical signal typically occupies only a small fraction of pixels. This becomes critical for the forthcoming CYGNO-04 demonstrator, where multiple high-resolution cameras at few-Hz rates yield sustained data throughput of $O(10^2)$ MB/s if full frames are stored. Traditional offline reconstruction pipelines [4] provide high-fidelity track characterization but are not suitable for trigger-level use, with per-frame latencies of order seconds. CYGNO therefore benefits from (i) *online localization and data reduction* via ROIs, and (ii) *NR/ER discrimination* even when event-level NR labels are not available.

In this contribution we focus on two ML strategies addressing these needs with minimal supervision: (1) a pedestal-trained autoencoder for unsupervised anomaly detection and fast ROI extraction, and (2) a weakly supervised CWoLa analysis using AmBe neutron calibration data to learn NR-like topologies from mixed samples.

2. Unsupervised ROI extraction from pedestal-trained anomaly detection

Pedestal data and preprocessing CYGNO camera frames are large, sparse, and dominated by structured noise (readout noise, fixed-pattern features, dark counts, and static optical artifacts). A distinctive asset of this detector is the availability of *pedestal frames* acquired with GEM amplification disabled, providing a high-purity sample of noise-only data for fully unsupervised training.

In this study, pedestal frames (training) are acquired with GEM voltages disabled, while standard frames (evaluation) are taken at nominal gain. To suppress edge instabilities observed in raw images [5], each frame is fiducialized by cropping to a central 1525×1525 region. Inputs are expressed relative to pedestal statistics using a consistent normalization chain and are downscaled to 1024×1024 for computational efficiency.

Autoencoder anomaly map and ROI construction Reconstruction-based anomaly detection with autoencoders (AEs) [6–8] is well matched to this setting. An AE trained exclusively on pedestal images learns to reconstruct the detector noise, while particle-induced structures present in standard frames appear as localized reconstruction failures. We study a convolutional AE with strided-convolution down-sampling, transposed-convolution up-sampling, skip connections between matching scales, and a 128-dimensional latent representation. For general background on these deep-learning components, see Ref. [6]. The anomaly map is defined as the pixel-wise residual $\mathbf{r}(\mathbf{x}) = |\mathbf{x} - \hat{\mathbf{x}}|$.

PoS(14YRM2025)003

A baseline configuration employs a hybrid Mean Square Error (MSE) plus Structural Similarity (SSIM) reconstruction objective [9]. Beyond this baseline, alternative training strategies are being studied to enhance sensitivity to faint, localized deviations from pedestal noise and improve anomaly-map quality, while limiting reconstruction of subtle structured features and preserving the learned noise morphology.

Residual maps are converted into ROIs using standard image-processing steps. A global threshold is applied (a representative value $\tau = 0.04$, chosen to suppress residual activity on held-out pedestal frames), followed by morphological closing to connect nearby residual fragments along individual tracks (using a circular structuring element with $d_{\text{link}} = 40$ pixels). The resulting binary mask defines the ROI. The selected working point lies on a broad plateau of the coverage–compression trade-off from a residual-threshold sweep, indicating stability against moderate threshold variations.

Performance The extracted ROIs are evaluated against the established offline reconstruction [4], which provides a high-fidelity physics reference but is not suitable for online use. Evaluation is performed on a per-event basis, requiring well-contained signals (at least one reconstructed pixel located more than 50 pixels from the image border). We report the signal-intensity coverage, the area cut defined as $f_{\text{cut}} = 1 - A_{\text{ROI}}/A_{\text{img}}$, and the per-frame inference latency.

On a test sample of 1563 reconstructed events, the current best-performing configuration achieves a signal-intensity coverage of $(93.0 \pm 0.2)\%$, an area cut of $f_{\text{cut}} = (97.8 \pm 0.1)\%$, and an inference latency of $t_{\text{inf}} \simeq 25$ ms per frame. These results illustrate the potential of pedestal-trained anomaly detection as a viable strategy for fast, unsupervised ROI extraction in CYGNO.

A broader discussion of this exploratory approach and of ongoing developments is presented in Ref. [10], currently submitted for publication to *Machine Learning: Science and Technology* (IOP) and available as an arXiv preprint.

3. Weakly supervised nuclear-recoil identification with CWoLa

AmBe mixed samples and mixture ceiling To obtain an NR-enriched sample without event-level labels, we exploit an Americium–Beryllium (AmBe) neutron source [11]. Neutrons are produced via the reactions $^{241}\text{Am} \to {}^{237}\text{Np} + \alpha$ and $^{9}\text{Be} + \alpha \to {}^{12}\text{C}^{*} + n$, with frequent de-excitation of $^{12}\text{C}^{*}$ emitting a 4.43 MeV γ ray. The emitted neutron spectrum extends up to ~ 11 MeV with a mean of ~ 4–5 MeV [11]. A dedicated GEANT4 study for a CYGNO-like configuration [12] indicates that neutrons contribute significantly to the AmBe-induced response in the active volume, producing NRs directly and additional ERs via associated γ production in surrounding materials. We define two datasets: an *AmBe* sample (source present, signal-enriched) and a *standard* (STD) sample (source absent, background-dominated), both restricted to a common region of the reconstructed energy–density plane (where the density is defined as the reconstructed energy divided by the number of hit pixels) in which AmBe shows a clear rate excess. In this region, a data-driven rate comparison yields a signal-attributable fraction $\alpha = (32.0 \pm 0.9)\%$. This mixture structure imposes a strict theoretical ceiling on ROC–AUC, $\text{AUC}_{\text{max}} = 0.5 + \alpha/2 = (0.660 \pm 0.005)$, since the irreducible $(1 - \alpha)$ background component in the AmBe data is indistinguishable from STD [13].

PoS(14YRM2025)003

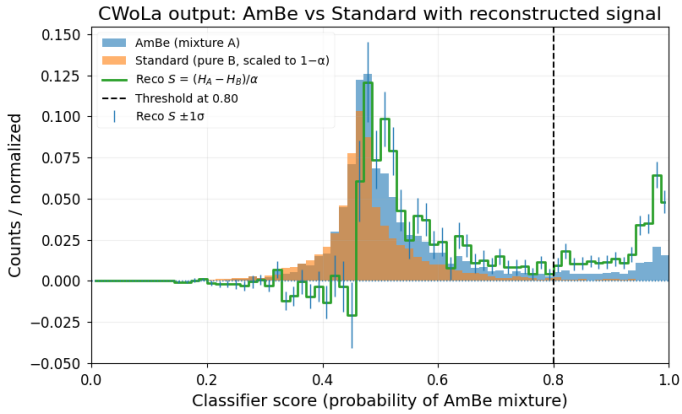

Figure 1: CWoLa classifier-score distributions for AmBe and standard test samples. The reconstructed signal component exhibits a peak near unity (NR-dominated) and an additional enhancement around ~ 0.5 (overlapping AmBe-induced populations). The dashed line indicates the NR-selection threshold.

Classifier and results The CWoLa approach [13] trains a classifier to distinguish the two mixtures using only sample-level labels; asymptotically this recovers the optimal S vs. B discriminator. We train a compact convolutional neural network on 128×128 grayscale images of reconstructed events, using the Adam optimizer [14] and binary cross-entropy while monitoring AUC. The test-set score distributions in Fig. 1 show an excess of high-score events in AmBe relative to STD, with a peak near unity (NR-dominated) and an enhancement around ~ 0.5 (overlapping source-induced populations). We select the NR-dominated tail with a conservative threshold $p > 0.8$.

Events above threshold populate the NR-favored region of the energy–density plane and show compact, approximately circular morphologies with high reconstructed density, consistent with nuclear recoils. These features serve as a qualitative consistency check, while quantitative validation with labeled simulations is deferred to future studies. The achieved performance approaches the theoretical ceiling, indicating near-optimal extraction of separation power from the mixed samples.

4. Conclusions

We summarized two complementary ML approaches for CYGNO with minimal supervision. A pedestal-trained anomaly-detection pipeline enables fast ROI selection from sparsely populated megapixel images, retaining $(93.0 \pm 0.2)\%$ of reconstructed signal intensity while discarding $(97.8\pm0.1)\%$ of image area at ~ 25 ms/frame, providing a realistic baseline for online data reduction. In parallel, a weakly supervised CWoLa classifier trained on mixed AmBe and standard samples learns NR-like topologies directly from data. A data-driven signal fraction $\alpha = (32.0 \pm 0.9)\%$ sets a theoretical AUC ceiling of (0.660 ± 0.005), and the observed performance approaches this limit while isolating compact, near-circular NR-like events. Together, these results support an ML-assisted path toward scalable online selection in next-generation CYGNO optical TPCs and, at the level of underlying methodology, more broadly applicable strategies for rare-signal localization and weakly supervised learning in other dark-matter searches.

PoS(14YRM2025)003

Acknowledgements This project has received fundings under the European Union's Horizon 2020 research and innovation program from the European Research Council (ERC) grant agreement No. 818744 and is supported by the Italian Ministry of Education, University and Research through the project PRIN: Progetti di Ricerca di Rilevante Interesse Nazionale "Zero Radioactivity in Future experiment" (Prot. 2017T54J9J). A. Messina has also been supported by the PNRR MUR project PE0000013–FAIR.

References

[1] Fernando Domingues Amaro et al. The CYGNO Experiment. *Instruments*, 6(1):6, 2022.

[2] Fernando Domingues Amaro et al. Bayesian network 3D event reconstruction in the Cygno optical TPC for dark matter direct detection. *Eur. Phys. J. C*, 85(11):1261, 2025.

[3] Fernando Domingues Amaro et al. The CYGNO experiment, a directional detector for direct Dark Matter searches. *Nucl. Instrum. Meth. A*, 1054:168325, 2023.

[4] F D Amaro et al. Directional idbscan to detect cosmic-ray tracks for the cygno experiment. *Measurement Science and Technology*, 34(12):125024, sep 2023.

[5] B. D. Almeida et al. Characterization of cutting-edge CMOS Active Pixel sensors within the CYGNO Experiment. 12 2025.

[6] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[7] Marco Farina, Yuichiro Nakai, and David Shih. Searching for New Physics with Deep Autoencoders. *Phys. Rev. D*, 101(7):075021, 2020.

[8] Tuhin S. Roy and Aravind H. Vijay. A robust anomaly finder based on autoencoders. 3 2019.

[9] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[10] F. D. Amaro et al. Fast reconstruction-based ROI triggering via anomaly detection in the CYGNO optical TPC. 12 2025.

[11] J.G. Fantidis. Comparison of different geometric configurations and materials for neutron radiography purposes based on a 241am/be neutron source. *Journal of Taibah University for Science*, 11(6):1214–1220, 2017.

[12] Flaminia Di Giambattista. *Background study of the LIME prototype at underground LNGS for the CYGNO experiment*. Phd thesis, Gran Sasso Science Institute, L'Aquila, Italy, March 2024. PhD in Astroparticle Physics, Cycle XXXV (2019–2023).

[13] Eric M. Metodiev, Benjamin Nachman, and Jesse Thaler. Classification without labels: Learning from mixed samples in high energy physics. *JHEP*, 10:174, 2017.

[14] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

PoS(14YRM2025)003