ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPM112

DETECTING ANOMALIES IN NON-STATIC ENVIRONMENTS: CONTINUAL LEARNING APPLIED TO CERN's KICKER MAGNET

M. Algelly*, F. Velotti, K. Papastegious, P. Ellison, V. Kain, F. Huhn, CERN, Meyrin, Switzerland
A. Kalousis, University of Geneva, Geneva, Switzerland

Abstract

This paper introduces an approach for unsupervised anomaly detection and early failure forecasting in CERN's Proton Synchrotron KFA71 extraction kicker system. We train a Conditional Variational Autoencoder (CVAE), conditioned on pulse settings, to reconstruct normal high-voltage waveforms and detect deviations. Forecasting anomalies involves monitoring a rolling mean of reconstruction error against a dynamic threshold derived from long-term statistics. To handle signal drift and evolving conditions, we use continual learning with Elastic Weight Consolidation and a replay buffer, showing adaptation without catastrophic forgetting.

INTRODUCTION

Beam extraction from the Proton Synchrotron accelerator to the TT20 transfer line relies on the KFA71 kicker system, composed of twelve independent high-voltage pulse generators. Since its commissioning in the 1970s, aging components (thyratrons, ferrites, transmission lines) and variable operational settings have led to intermittent failures and increased downtime. Traditional supervision methods, such as manual waveform inspections and threshold-based alarms, are brittle and sensitive to noise and drift.

Legacy methods also struggle with the volume and variability of waveform data, while fault rarity limits supervised learning. Recently, autoencoder-based anomaly detection has shown promise in CERN systems like the SPS Beam Dump System (SBDS) [1], and CMS and ATLAS experiments [2–4], proving robust in complex environments.

To address KFA71 challenges, we propose an unsupervised method using reconstruction error from a Conditional Variational Autoencoder (CVAE), conditioned on pulse settings. This work adds a forecasting strategy based on reconstruction trends and tackles non-stationarity via continual learning with Elastic Weight Consolidation and replay. We demonstrate that this combination enables robust detection and early warning under evolving conditions.

SYSTEM AND DATA

Hardware and Acquisition

A typical cycle of the system starts with the Resonant Charging Power Supply (RCPS) charging the Pulse Forming Line (PFL). On trigger, the main switch discharges the PFL into a transmission line, generating an 80 kV pulse toward the magnet. Residual energy is dissipated via a dump switch and resistor. The pulse deflects the beam, with reflections absorbed by a terminating resistor. A overview of the kicker module architecture can be found in [5].

Dataset

A digitized version of the voltage measured at the terminating resistor, hereafter referred to as a waveform, is recorded for each pulse, sampled every 4 ns over a 10 µs interval (2,500 points per waveform), as shown in Fig. 1. The dataset includes several million such waveforms collected over October–November 2024, each accompanied by pulse settings (target voltage and duration). This period was selected in consultation with system experts as it provides a large and representative sample of real operational conditions for training and evaluation.

A notable challenge is the significant imbalance in pulse settings: roughly 98% of pulses use only two lengths, and 80% focus around six voltage values. This imbalance risks overfitting dominant configurations, potentially causing false anomalies in rare settings. To mitigate this, we apply balancing methods, including sampling strategies, and targeted evaluation on underrepresented configurations.

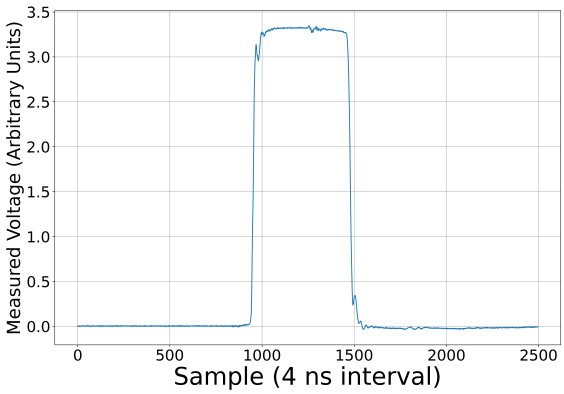

Figure 1: Voltage measured at the terminating resistor.

ANOMALY DETECTION METHOD

Our method trains a generative model on nominal data, using reconstruction error as an anomaly score. We employ a Conditional Variational Autoencoder (CVAE) [6], an unsupervised architecture with an encoder, conditioning path, and decoder, as shown in Fig. 2.

The *WaveformEncoder* maps each input waveform $x \in \mathbb{R}^N$ (with $N = 2500$) to the parameters $(\mu_{wave}, \text{logvar}_{wave})$ of a Gaussian distribution. The conditioning variable $p \in \mathbb{R}^2$ represents the pulse settings: target voltage and pulse length. The CVAE loss combines a mean-squared-error

* malik.algelly@cern.ch

ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPM112

(MSE) reconstruction term and a Kullback–Leibler (KL) regulariser:

$$\mathscr{L}(x) = \underbrace{\mathbb{E}_{q(z|x,p)}[\|x - \hat{x}\|^2]}_{\mathrm{MSE}(x,\hat{x})} + \beta_{KL} D_{\mathrm{KL}}(q(z|x,p) \,\|\, p(z)) \tag{1}$$

where $\hat{x}$ is the reconstructed waveform, $p(z)$ is the prior distribution, $q(z|x,p)$ is the variational posterior, and β_{KL} is a weighting factor. The MSE term forces $\hat{x}$ to match x point-wise for nominal pulses, so large values of $r = \|x - \hat{x}\|^2$ naturally flag anomalies. The KL divergence $D_{\mathrm{KL}}(q, ||, p)$ measures how much the learned distribution q diverges from the prior p, and penalizes deviations from the desired latent structure [7].

As discussed in the introduction, this class of autoencoder-based methods has already demonstrated strong performance in other CERN systems, including the SPS Beam Dump System and LHC experiments [1–4].

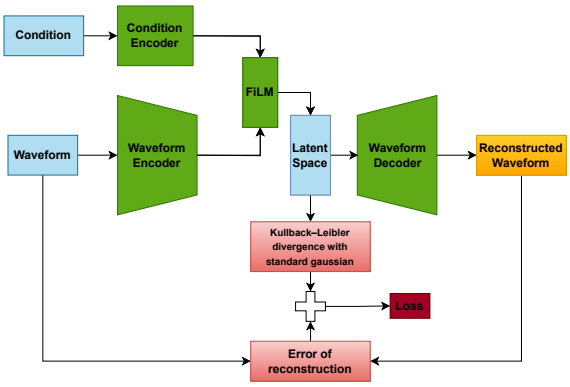

Figure 2: Conceptual overview of the Conditional VAE architecture.

Conditioning is incorporated using a FiLM-like mechanism [8]. Pulse settings p are processed by the *ConditionEncoder* and associated FiLM layers to generate scaling vectors $\vec{\gamma}$ and shifting vectors $\vec{\beta}$. These vectors then modulate the initial latent parameters $(\mu_{wave}, \mathrm{logvar}_{wave})$ from the *WaveformEncoder* to produce the final conditioned parameters (μ, logvar) for $q(z|x,p)$. A latent vector z is subsequently sampled from this conditioned Gaussian distribution $\mathcal{N}(\mu, \mathrm{diag}(\exp(\mathrm{logvar})))$ and passed to the *WaveformDecoder* to reconstruct the waveform $\hat{x}$.

After training, the model computes reconstruction error $r = \|x - \hat{x}\|^2$ per waveform. High errors typically signal deviations from the training distribution. To validate this, we manually labeled notable anomalies, forming a preliminary ground truth for clear faults. Figure. 3 shows reconstruction errors over time, with labeled anomalies (in red) often aligning with major spikes—highlighting the model's effectiveness. Notably, the model also flags additional pulses not identified during manual inspection; further analysis confirmed that many of these are valid anomalies that were initially missed.

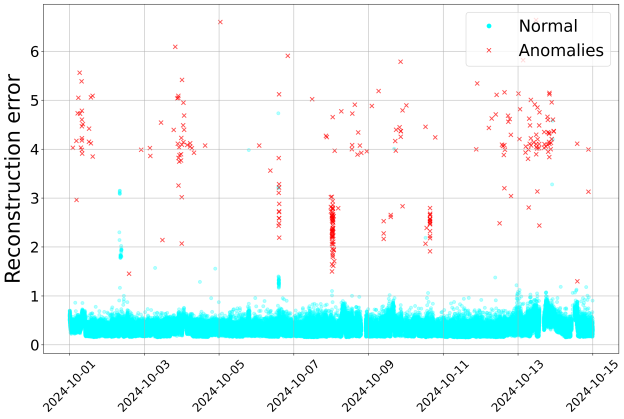

Figure 3: Reconstruction error per pulse. Manually labeled anomalies are highlighted in red.

ANOMALY FORECASTING

While anomaly detection identifies individual faulty pulses, anticipating failures ahead of time is essential for operational mitigation. Empirically, we observed that reconstruction error $r = |x - \hat{x}|^2$ often increases gradually in the minutes preceding a failure. This motivated the design of an early-warning mechanism based on the temporal evolution of reconstruction errors.

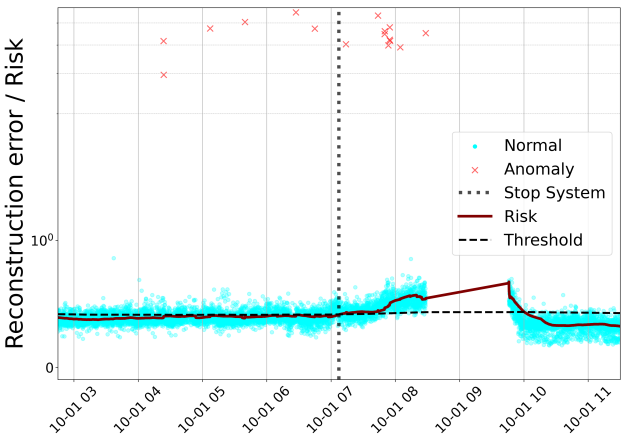

Figure 4: Forecasting based on risk score and dynamic threshold. Vertical dashed lines indicate warning triggers.

To quantify this evolution, we compute a risk score at each time point as the rolling average of reconstruction errors over the past 30 minutes. To assess whether the current risk level is abnormal, we compare it to a dynamic threshold defined as the median of the risk score over the last 12 hours, augmented by a multiple of its median absolute deviation (MAD). The sensitivity of the detection is controlled by a tunable scaling factor.

The 30-minute averaging smooths short-term fluctuations while capturing gradual trends, and the 12-hour baseline ensures robustness to long-term variations in beam settings or system behaviour. An early warning is raised when the

ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPM112

risk score remains above this threshold for more than five minutes, providing a practical signal of upcoming instability.

Despite its simplicity, this approach effectively identifies periods of degradation well in advance of hard failures, offering a first step toward reliable forecasting using reconstruction dynamics. An example of this forecasting method applied to real data is shown in Fig. 4.

CONTINUAL LEARNING

The waveform distribution evolves over time due to a variety of factors, including hardware upgrades, shifts in pulse tuning, and gradual changes in environmental or system conditions such as temperature. These changes can cause consistent drifts in waveform shape. These shifts can severely degrade the performance of a static model trained on past data. To address this, we adopt Continual Learning techniques that allow the model to adapt while preserving prior knowledge. We implement Elastic Weight Consolidation (EWC) [9], which adds a regularisation term in the loss function to prevent critical parameters—identified using the Fisher Information Matrix—from drifting too far.

The modified training loss becomes:

$$\mathscr{L}_{\mathrm{EWC}}(\theta) = \mathscr{L}_{\mathrm{VAE}}(\theta) + \lambda \sum_i F_i(\theta_i - \theta_i^*)^2 \qquad (2)$$

where θ are the current model parameters, θ_i^* are the parameters learned on previous data, λ controls the strength of regularisation, and F_i estimates the importance of parameter θ_i. This importance is computed from the diagonal of the Fisher Information Matrix, which measures how sensitive the model likelihood is to changes in each parameter:

$$F_{ij} = \mathbb{E}_{x\sim p(x|\theta)}\left[\frac{\partial \log p(x|\theta)}{\partial \theta_i}\frac{\partial \log p(x|\theta)}{\partial \theta_j}\right] \qquad (3)$$

In practice, only the diagonal terms $F_i = F_{ii}$ are retained, and the expectation is approximated by an empirical average over the dataset:

$$F_i \approx \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\partial \log p(x_n|\theta)}{\partial \theta_i}\right)^2 \qquad (4)$$

As shown in [10], under the assumption of a Gaussian decoder with fixed variance, these gradients are proportional to those of the MSE loss, making the Fisher terms efficiently computable during standard training.

However, EWC alone suffers from a stability–plasticity trade-off: large λ values preserve old knowledge but block adaptation, while small λ enable learning but cause forgetting. This is illustrated in Fig. 5. To mitigate this, we introduce a replay buffer containing a small amount of past waveforms, which are interleaved with new data during updates. This hybrid EWC + Replay strategy [11] maintains adaptation while preventing catastrophic forgetting.

Figure 6 illustrates the impact of our continual learning strategy. The initial model (”Before update”, red dots) shows increased reconstruction error during a drift period starting

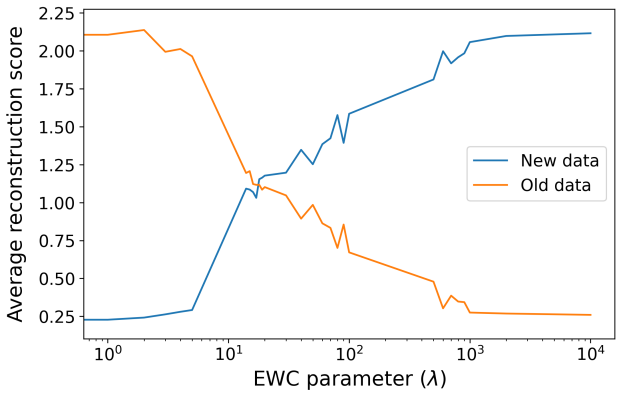

Figure 5: Stability–plasticity trade-off with EWC.

in mid-October, with some recovery afterward. In contrast, the model updated using EWC + Replay (”After update”, blue dots) maintains consistently low error. It successfully adapts to the drift, handles the return to nominal conditions, and crucially preserves performance on earlier October data. This demonstrates the method’s ability to adapt to evolving distributions while avoiding catastrophic forgetting, enabling robust operation in non-stationary environments.

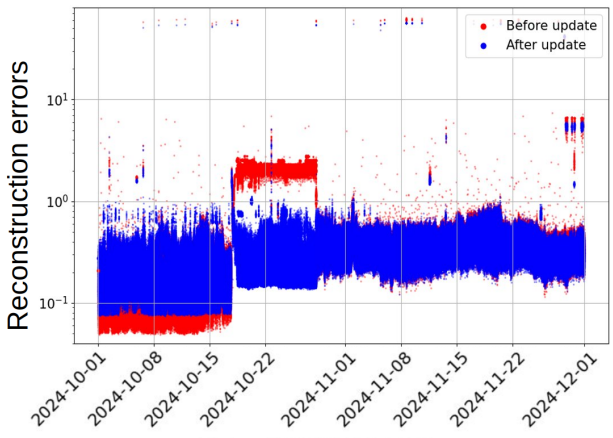

Figure 6: Impact of continual learning: comparison of reconstruction errors before and after model update.

CONCLUSION AND PERSPECTIVES

We demonstrated that a FiLM-conditioned CVAE models the diverse KFA71 waveforms, detects anomalies, and forecasts failures. Continual learning with EWC and a small replay buffer preserves accuracy as conditions drift.

Next steps will focus on rare pulse modes, add ARIMA/LSTM [12, 13] forecasters with uncertainty estimates, and test longer-horizon continual-learning schemes such as variational methods [14]. Operational deployment needs a light MLOps stack: versioned models and datasets, real-time performance/drift dashboards with one-click rollback, and an operator-friendly view for threshold tuning and system status.

ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPM112

REFERENCES

[1] F. Huhn, B. Goddard, F. Velotti, and V. Bencini, "Automating beam dump failure detection using computer vision", in *Proc. IPAC'23*, Venice, Italy, May 2023, pp. 4456–4459. doi:10.18429/JACoW-IPAC2023-THPL017

[2] CMS ECAL Collaboration, "Autoencoder-based anomaly detection system for online data quality monitoring of the CMS Electromagnetic Calorimeter", *Comput. Softw. Big Sci.*, vol. 8, 11, 2024. doi:10.1007/s41781-024-00118-z

[3] A. Gandrakota, "Real-time anomaly detection at the L1 trigger of CMS experiment", *arXiv:2411.19506, 2024*. doi:10.48550/arXiv.2411.19506

[4] ATLAS Collaboration, "Search for new phenomena in two-body invariant mass distributions using unsupervised machine learning for anomaly detection at $\sqrt{s} = 13$ TeV", *Phys. Rev. Lett.*, vol. 132, no. 8, 081801, 2024. doi:10.1103/PhysRevLett.132.081801

[5] M.J. Barnes *et al.*, "Kicker systems - part 1 - introduction and hardware", presented at the CERN Accelerator School (CAS), March 2017. https://indico.cern.ch/event/451905/contributions/2159055/attachments/1424705/2515898/BARNES_CAS_Mar17_Kickers-PtA.pdf

[6] A.A. Pol, V. Berger, G. Cerminara, C. Germain, and M. Pierini, "Anomaly detection with conditional variational autoencoders", *arXiv:2010.05531, 2020*. doi:10.48550/arxiv.2010.05531

[7] S. Kullback and R.A. Leibler, "On information and sufficiency", *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951. doi:10.1214/aoms/1177729694

[8] E. Perez, F. Strub, H. de Vries, V. Dumoulin, and A. Courville, "FiLM: Visual reasoning with a general conditioning layer", *arXiv:1709.07871, 2017*. doi:10.48550/arxiv.1709.07871

[9] J. Kirkpatrick *et al.*, "Overcoming catastrophic forgetting in neural networks", *arXiv:1612.00796, 2017*. doi://10.48550/arxiv.1612.00796

[10] O. Rybkin, K. Daniilidis, and S. Levine, "Simple and effective VAE training with calibrated decoders", *arXiv:2006.13202, 2021*. doi:10.48550/arxiv.2006.13202

[11] K. Faber, R. Corizzo, B. Śnieżyński, and N. Japkowicz, "VLAD: Task-agnostic VAE-based lifelong anomaly detection", *Neural Networks*, vol. 165, pp. 248-–273, Aug. 2023. doi:10.1016/j.neunet.2023.05.032

[12] G.E.P. Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, 1970.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. doi:10.1162/neco.1997.9.8.1735

[14] C.V. Nguyen, Y. Li, T.D. Bui, and R.E. Turner, "Variational continual learning", *arXiv:1710.10628, 2018*. doi:10.48550/arxiv.1710.10628