UNSUPERVISED ANOMALY DETECTION AND CHANNEL ATTRIBUTION WITH VARIATIONAL AUTOENCODERS AT THE ADVANCED LIGHT SOURCE

A. Sulc*, T. Hellert, S. C. Leemann, G. Martino, H. Nishimura
Lawrence Berkeley National Laboratory, Berkeley, CA, U.S.A.

Abstract

We present an unsupervised pipeline that learns a compact representation of beam-on machine state at the ALS, detects anomalies preceding beam-loss events, and highlights the responsible channels for operator diagnosis. Archiver data are resampled to a uniform time grid, filtered to beam-on intervals using stored current, and pruned by variability and principal-component analysis. A variational autoencoder with residual encoder-decoder stacks is trained on the standardised PV vectors; the global anomaly score and the per-PV attribution are both derived from per-PV reconstruction z-scores, so the score is an exact decomposition of the channel ranking. We apply the pipeline to 34 beam-loss events from the 2025 ALS user run; in several cases it surfaces early-stage anomalies in the PV subsystems that subsequently led to the beam dump, indicating a framework can act as an early-warning aid for operators.

INTRODUCTION

At third-generation light sources such as the Advanced Light Source (ALS) [1], every unplanned beam dump erodes delivered user beamtime, and recovery hinges on identifying precursors hidden in the minutes before the dump. Those precursors live in a high-dimensional, heterogeneous EPICS [2] stream—magnet power supplies, vacuum gauges, RF cavities, interlocks, cooling circuits—updated continuously and far exceeding what an operator can monitor channel-by-channel in real time.

Supervised classifiers are impractical in this setting: labelled failures are sparse, fault categories evolve, and non-stationary drift across months of operation invalidates static decision boundaries. We instead train a variational autoencoder directly on raw beam-on archiver data, with no external labels and no time axis as a feature: the encoder treats each timestep as an independent PV vector, so what it learns is the hidden cross-channel correlation structure that distinguishes one operating mode from another. Departures from that learned structure both raise a global alarm and surface the responsible channels, while the latent space itself doubles as an differentiable, low-dimensional map of machine state.

Variational autoencoders (VAEs) [3] offer a principled unsupervised alternative. By maximising a tractable evidence lower bound on the log-likelihood of observed proces variable (PV) vectors, the VAE learns a low-dimensional latent manifold of nominal beam-on machine state. Channels whose reconstruction departs from the model's nominal expectation can then be flagged individually. The global anomaly score is built from those same per-channel deviations, so the alarm signal and the per-PV ranking are not separate post-hoc artefacts.

This contribution describes the offline toolchain developed for ALS fault analysis: archiver data alignment and beam-on filtering; channel curation via variability and PCA scoring; the VAE architecture and training protocol; and per-fault anomaly scoring, embedding visualisation, and per-PV reconstruction-residual ranking, all derived from the same per-channel calibration.

* asulc@lbl.gov

DATA PIPELINE

Archiver Retrieval and Time Alignment

The model was trained from manually curated lists of variables. All PVs contributing to a model share a uniform time grid spanning the global data range with configurable step Δt. For training the facility-wide model we use $\Delta t = 1800s$ (30 min), matching the typical PV update cadence of slow diagnostic channels; for per-fault analysis the grid is refined to $\Delta t = 10s$ to resolve faster transients. Values falling within a bin are aggregated by arithmetic mean. For the stored-current PV (`SR:DCCT`) that gates beam-on selection, aggregation uses the bin maximum so that brief excursions above threshold are not missed. Bins without samples are forward-filled (sample-and-hold), consistent with the control-system semantics in which a PV retains its last published value until the next update.

Beam-On Filtering and Channel Smoothing

Training and baseline statistics use only timesteps with `SR:DCCT` above 1 mA (and > 50 mA for fault-window display), excluding injection gaps and empty-ring periods. Near-discrete PVs (≤ 10 unique levels) are label-smoothed to $[\varepsilon, 1-\varepsilon]$ with $\varepsilon = 0.05$ so mean squared error (MSE) gradients remain informative for these channels.

CHANNEL CURATION

Before VAE training, PVs are filtered to retain only informative, non-redundant signals. The curation pipeline ranks channels by variance, range, "active fraction" (fraction of timesteps where $|x - \text{median}| > 0.5\,\sigma$), and coefficient of variation. PVs with fewer than 4096 raw archiver samples or with near-zero variance ($\sigma^2 < 10^{-12}$) are removed unconditionally. An optional bottom-percentile cut on raw variance removes the flattest channels.

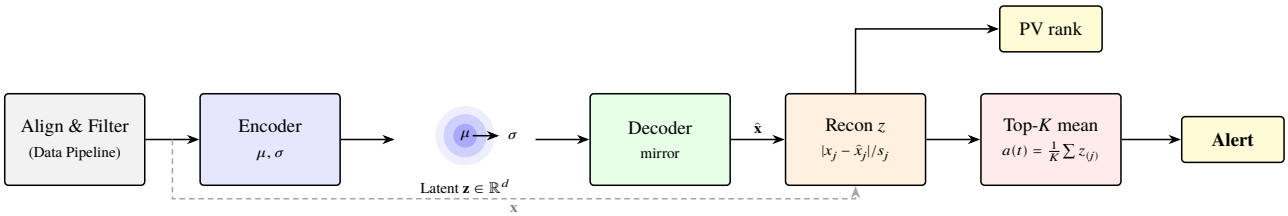

Figure 1: Anomaly detection pipeline. Per-PV residual z-scores computed from the decoder output drive both the global score $a(t)$ and the per-PV ranking, making score and attribution exact decompositions of the same $z_j(t)$.

After per-PV standardisation, PCA is fitted with up to 50 components and an importance score is computed as the sum of squared loadings on the leading $K = 20$ components. The resulting exclude list is consumed by the training script, keeping the trained model focused on informative dynamics; the model used for the results below retains $D = 1497$ PVs after curation.

VAE ARCHITECTURE AND TRAINING

This section details the network architecture and training protocol; a conceptual overview of the anomaly detection pipeline is illustrated in Figure 1.

Network Design

The encoder comprises a stack of fully connected layers with batch normalisation, ReLU activations, and dropout ($p = 0.1$), each followed by L pre-activation residual blocks [4] of matching width. Two linear heads emit the latent mean $\mu \in \mathbb{R}^d$ and log-variance $\log \sigma^2 \in \mathbb{R}^d$ for a diagonal Gaussian posterior $q_\phi(\mathbf{z}\,|\,\mathbf{x}) = \mathcal{N}(\mu, \mathrm{diag}(\exp(\log \sigma^2)))$. The decoder mirrors the encoder and maps $\mathbf{z}$ back to $\mathbb{R}^D$ (the PV count). Initialisation sets final decoder weights and biases to zero and biases $\log \sigma^2$ to -2 in the encoder, so early training emphasises reconstruction before the KL term becomes significant.

Objective

Inputs are standardised per PV (zero mean, unit variance with a variance floor of 10^{-8}). The loss combines mean-squared reconstruction error with the KL divergence from the standard normal prior $\mathcal{N}(\mathbf{0}, \mathbf{I})$, weighted by β:

$$\mathcal{L} = \mathrm{MSE}(\hat{\mathbf{x}}, \mathbf{x}) + \beta\, \mathrm{KL}(q_\phi(\mathbf{z}|\mathbf{x}) \,\|\, p(\mathbf{z})). \tag{1}$$

Following the β-VAE framework [5], β is linearly ramped from 0 to its target over the first $N_w = 50$ warm-up epochs to mitigate posterior collapse. Optimisation uses Adam (learning rate 10^{-3}, batch size 1024, 128 epochs) with gradient clipping (max norm 1.0). The training cadence is $\Delta t = 1800$ s; the per-fault analysis cadence is $\Delta t = 10$ s. The trained model used for the results below has latent dimension $d = 20$, hidden widths $(512, 256, 128)$, $L = 2$ residual blocks per width, dropout 0.10, and retains $D = 1497$ PVs after curation.

Checkpoint Contents

The saved model stores input dimension, normalisation tensors (per-PV mean and standard deviation), PV ordering, and all architectural hyperparameters, so downstream fault scripts can reconstruct the identical preprocessing pipeline without external configuration.

FAULT-WINDOW ANALYSIS

For each beam-loss event in the ALS operational log, archiver data are downloaded over a 48-hour UTC window (the calendar day prior through the end of the fault day) and aligned at $\Delta t = 10$ s with the same PV ordering as the trained model.

Per-PV Reconstruction z-Score

We use the deterministic VAE reconstruction $\hat{\mathbf{x}} = \mathrm{decode}(\mu(\mathbf{x}))$. Per-PV residuals $r_j(t) = x_j(t) - \hat{x}_j(t)$ are evaluated in standardised input units, and their per-PV baseline statistics, median $\tilde{r}_j$ and robust scale $s_j = \mathrm{MAD}_j$ (mean absolute deviation), floored at 0.01 to suppress near-constant baseline channels, are estimated from a beam-on window of H_b hours before the fault for which the stored current exceeds $I_{\min}$ (defaults $H_b = 4$ h, $I_{\min} = 50$ mA). The calibrated per-PV z-score is

$$z_j(t) = \frac{|r_j(t) - \tilde{r}_j|}{s_j}, \tag{2}$$

which yields a more scale-invariant deviation.

Top-K Anomaly Aggregation

The global anomaly score is the top-K mean of the per-PV z-scores at each instant, $a(t) = \frac{1}{K}\sum_{j \in \text{top-}K(t)} z_j(t)$, with $K = 10$ in the experiments below. Because the score and the per-PV ranking are built from the same $\{z_j(t)\}$, the alarm at instant t and the channel attribution are exact decompositions of each other: the score $a(t)$ is, by construction, the average of the K PVs the ranking surfaces. A causal moving average (≈ 2.5 min) smooths $a(t)$ without future lookahead. The alarm threshold is $\bar{a}_b + 3\, s_{a,b}$, where $\bar{a}_b$ and $s_{a,b}$ are the mean and standard deviation of the raw aggregated score $a(t)$ over the same beam-on baseline window used to calibrate the per-PV residual statistics, and the earliest connected above-threshold run before the fault defines the precursor warning time.

Reconstruction-Residual Channel Ranking

PVs are ranked by the mean of z_j over the anomalous interval (the connected above-threshold run on the smoothed $a(t)$, or the trailing 10 min when no warning is found) and

overlaid against the beam-current trace, directing operators to physically interpretable candidates. Because the same z_j that feed the score determine this ranking, the channels listed in the attribution are by construction those responsible for the value of $a(t)$ at the alarm.

RESULTS

The pipeline was applied to a set of beam-loss events from the 2025 ALS user runs, covering categories including power-supply trips, vacuum excursions, RF faults, water-flow interruptions, and control-system failures. We do not have labelled ground-truth precursor timestamps for these events, so rather than report a single detection-rate number, we present a qualitative, figure-driven assessment of behaviour on both training (nominal) data and faults.

Figures 2 and 3 show two illustrative per-fault windows, each combining the smoothed top-K mean z-score trace (top panel) with the per-PV z-score time series of the highest-ranked channels (bottom panel). The first fault (#3246, Fig. 2) is a storage-ring bend magnet power supply trip that the operator logbook attributes to an SCR over-temperature and is resolved by placement of a flow control unit (cooling water). The aggregated trace crosses the beam-on 3σ threshold a short time before the dump, giving a precursor of order ~15 min.

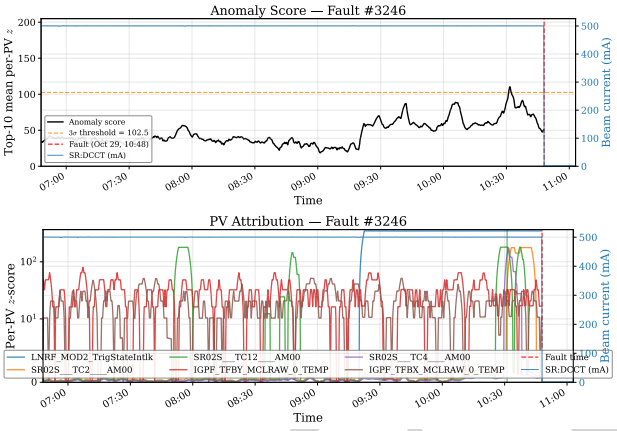

Figure 2: Fault #3246 (SR B trip, SCR over-temp, H_2O system). *Top:* smoothed top-K mean per-PV z-score in the 4 h preceding the fault; the trace exceeds the beam-on 3σ threshold (dashed) before beam loss. Right axis: SR:DCCT (mA); the trace is masked where $I < 50$ mA. *Bottom:* per-PV z-score traces for the highest-ranked channels in the attribution window.

The second event (#3242, Fig. 3) corresponds to a logbook entry noting that an LFB amplifier failed and was swapped. For this event the per-PV ranking is led by transverse-feedback amplifier diagnostics (`IGPF_TFBY_MCLRAW_1_REV/TEMP/FWD`) together with the LFB system's bunch oscillation RMS readback (`IGPF_LFB_SRAM_MAXRMSVAL`). These channels could plausibly be associated with a failing feedback amplifier, a failing amplifier would presumably disturb its own diagnostics, but we report the correspondence as suggestive rather than definitive, since the ranking alone does not identify which of these channels is causal.

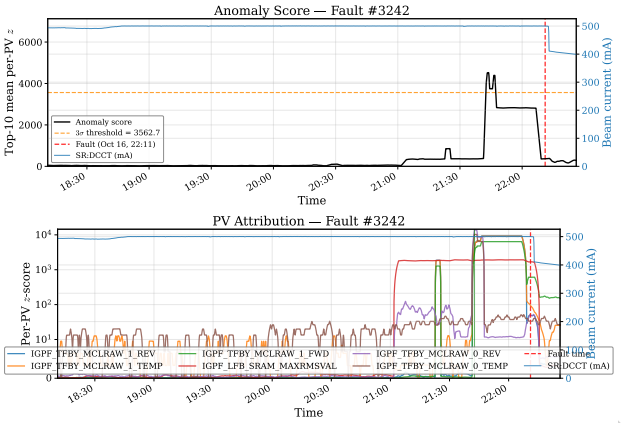

Figure 3: Fault #3242 (LFB amplifier swap, LFB-THC system), with the same layout as Fig. 2.

A quantitative detection-rate or lead-time estimate would require a separate, labelled validation set with operator-confirmed precursor timestamps, which is left to future work.

CONCLUSION

We presented a reproducible, unsupervised pipeline from EPICS archiver exports to VAE-based anomaly scores built as a top-K aggregation of per-PV reconstruction z-scores, per-fault t-SNE embedding visualisations, and a residual-based channel ranking derived from the same per-PV z-scores. The method requires no labelled fault data and adapts to evolving machine conditions through periodic retraining on beam-on archiver snapshots. Preliminary results suggest that this aggregated z-score trace can provide an interpretable precursor signal in advance of a subset of beam-loss events, and that the associated per-PV ranking tends to direct attention to physically related PV subsets.

ACKNOWLEDGEMENTS

This work was supported by the Director of the Office of Science of the U.S.Department of Energy under Contract No. DEAC02-05CH11231.

REFERENCES

[1] T. Hellert *et al.*, "Status of the Advanced Light Source", in *Proc. IPAC'24*, Nashville, TN, USA, pp. 1309–1312, May 2024. doi:10.18429/JACoW-IPAC2024-TUPG37

[2] L. R. Dalesio *et al.*, "The experimental physics and industrial control system architecture: past, present, and future", *Nucl. Instrum. Methods Phys. Res. A*, vol. 352, no. 1, pp. 179–184, 1994. doi:10.1016/0168-9002(94)91493-1

[3] D. P. Kingma and M. Welling, "Auto-encoding variational bayes", Dec. 2013. doi:10.48550/arXiv.1312.6114

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Jun. 2016. doi:10.1109/CVPR.2016.90

[5] I. Higgins *et al.*, "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework", in *Proc. ICLR'17*, Toulon, France, Apr. 2017.