Beam-Based Anomaly Detection in Accelerators Using Variational Autoencoders with Drift-Aware Continual Learning

Edoardo Barbi

September 29 2025

Introduction

FACET-II

- Located at SLAC National Accelerator Laboratory, it uses the middle kilometer of the original 3 Km linac.
- It produces beams of highly energetic electrons (up to 10 GeV) with extremely high peak currents.
- Its primary mission is to research advanced accelerator technologies, with a focus on plasma wakefield acceleration (PWFA).

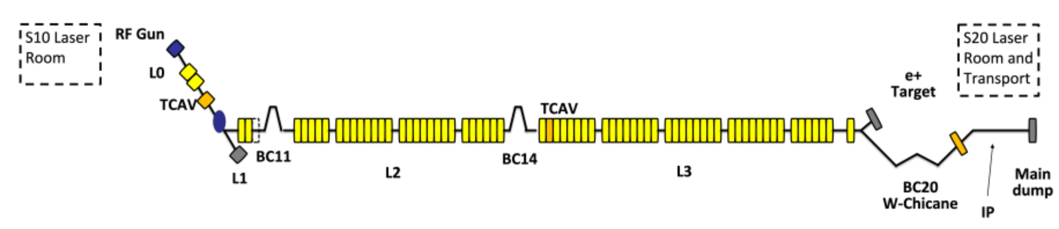


Figure 1: Schematic layout of the FACET-II facility.

FACET-II Control System

- FACET-II uses the EPICS software for its control system.
- EPICS assigns each component of the accelerator to a Process Variable (PV), for example `BPMS:IN10:731:X` for the X orbit position measured by the Injector's beam position monitor.
- FACET-II has 3 types of PVs:
  - BSA (beam synchronous) for important PVs such as BPMS in the injector.
  - SCP, part of the older SLAC control system.
  - Non-BSA for relatively static PVs (such as QUADS values).

Thesis Goal

Goal: develop a system that can **continuously** flag anomalies in the FACET-II's injector (IN10) and Sector 11 (LI11), using only BSA PVs.

A particle accelerator is a system that is subject to **drift** (e.g. change in temperature over the day) $\rightarrow$ the model will also have to **adapt itself** as it goes through the data based on how much drift is detected.

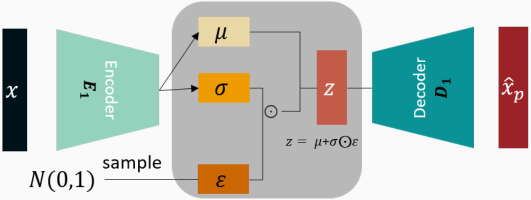
Common Types of Accelerator Anomalies

- **RF System Instabilities:** Crucial for acceleration, but prone to faults.
  - **Amplitude Jumps:** Sudden, discrete changes in a klystron's output power, which directly impacts beam energy.
  - **Amplitude Jitter:** High-frequency oscillations in RF power, leading to shot-to-shot instability in the beam.
- **Beam Trajectory Deviations (Orbit Drifts):** The path of the beam is not static.
  - **Orbit Jumps:** Abrupt shifts in the beam's position. This is often a symptom of an upstream issue, like a timing or RF phase fluctuation.
- **Collective Effects:** The beam's own fields causing self-interaction.
  - **Transverse Wakefields:** An off-axis beam can induce fields in the beam pipe that deflect the tail of the bunch, leading to emittance growth.

Variational Autoencoders

Variational Autoencoder (VAE): The Basics

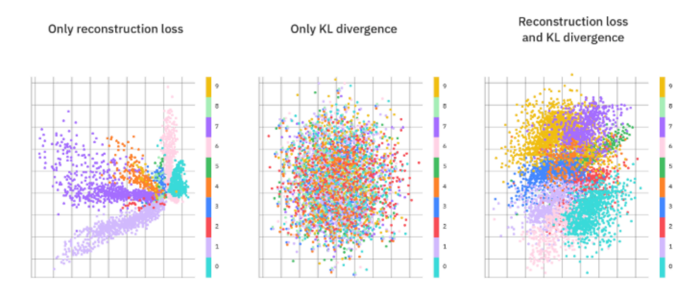
- An **unsupervised** model learning a probabilistic latent representation.
- **Encoder**: maps input ($\mathbf{x}$) to the parameters (μ, σ) of a latent distribution.
- **Latent vector** ($\mathbf{z}$) is sampled from this distribution: $\mathbf{z} = \mu + \sigma \odot \varepsilon$.
- **Decoder**: reconstructs the input ($\hat{\mathbf{x}}_p$) from the sampled $\mathbf{z}$.



VAE Architecture

Why a VAE

- VAE vs standard AE:
  - VAEs have an extra term in the loss function (**KL divergence**).
  - $\mathscr{L}_{VAE} = loss + KL$
  - $\mathscr{L}_{AE} = loss$



Latent space visualization with different loss components.

Anomaly detection

Anomaly detection

The Key Idea: Train an ensemble (5) of variational autoencoders (mixture of experts) **only** on data from normal, stable operation using PVs from the Injector and Sector 11.

- The models become experts at reconstructing ”normal” machine states.
- When an anomalous state is given as input, the models struggle, and the reconstruction is poor.

The **Reconstruction Error** is calculated as the average between the 5 models and used as our anomaly score.

$$\text{Anomaly Score} = ||\text{Input} - \text{Reconstructed Output}||^2$$

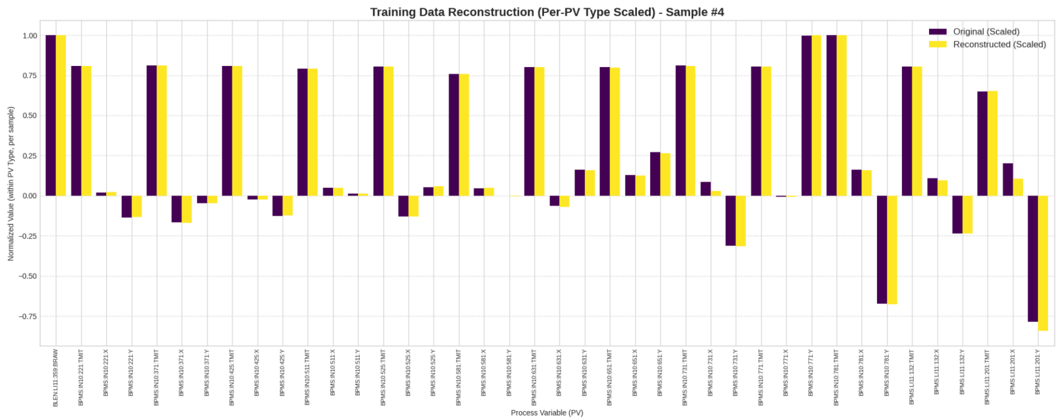
A high score indicates a likely anomaly.

Training the VAEs

The models were trained on data from the Injector and the first sector after that, applying **stable beam filters** on the following PVs:

- LASER
- BPMs
- BLENs
- KLYS (amplitude and phase)
- TOROIDS

VAE reconstruction on training data



The VAE demonstrates strong reconstructive capabilities. Slight inaccuracies (within a few percent) are tolerable as the aim is not perfect data prediction, but the detection of substantial reconstruction failures.

After Training: Applying the Model

To ”simulate” the flow of data as if the model was used on real time accelerator data, the following approach was used:

1. Take approx. 2 days of relatively stable data
2. **Train** the model on 10 hours of data with stable beam filters
3. Take the subsequent data after training and present it to the model in 2 minute windows, and look for anomalies
4. The system has to fine-tune itself in every window to avoid becoming obsolete

The Challenge: Concept Drift

A model trained at one point in time will not be reliable forever.

The accelerator’s ”normal” state slowly changes over time due to **drift**.

As a result, the model’s reconstruction error for ”normal” data will slowly increase, leading to false alarms.

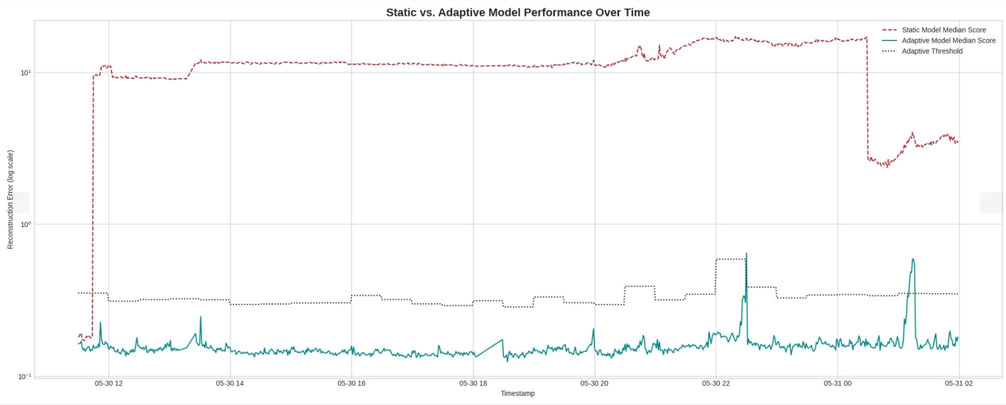
Solution: Drift-Aware Continual Learning

To keep the VAE model reliable, we make it adaptive.

The Process:

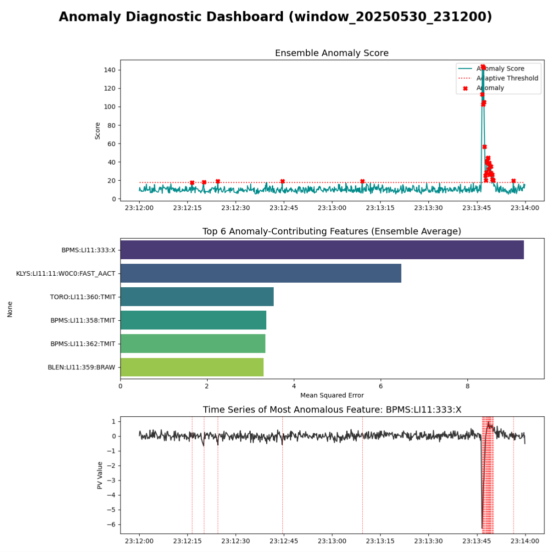
1. **Detect Drift:** Continuously monitor the **stable** input data and model performance using statistical tests: KS-test, MWU-test for distribution drift and Mahalanobis distance for performance drift. Combined they produce a **drift score** ($\varepsilon[0,1]$).
2. **Trigger Adaptation:** When drift is detected, automatically trigger a model update cycle.
3. **Adapt the Model:** Fine-tune the VAE on the stable data of the current window. Including a small amount of old data (**experience replay**) prevents the model from forgetting past states. The extent of the retraining (learning rate and epochs) depend on the drift score.

Static VS Adaptive model



A **”static model”** trained only on the initial stable beam data and applied to subsequent windows becomes ”outdated” almost immediately, flagging most of the new windows data as anomalies, while the **adaptive model**, keeps a consistent anomaly scoring over time

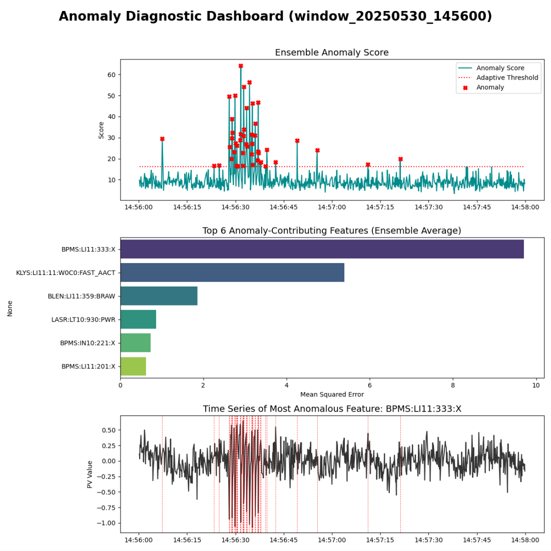
Example Anomaly Plot (2 min window)



Clear anomaly example: the energy BPM had a severe drop in value.

- The model, correctly detects the anomaly and returns the most anomalous features.
- The anomaly was given by a jump in the Klystron’s amplitude, correctly flagged.

Example Anomaly Plot



- High-frequency amplitude jitter
- Originates from **KLYS:LI11:11:W0C0:FAST_AACT**
- Propagates to the downstream **BPM**
- Model detects:
  - **Cause:** jittering RF source
  - **Effect:** unstable beam energy proxy
- Shows robustness to both **single-shot jumps** and **periods of high variance**

Limitations

- The model can only detect shot to shot anomalies (no time dependent ones)
- The model assumes that if the underlying variables distributions change over time, it’s due to drift and not an anomaly

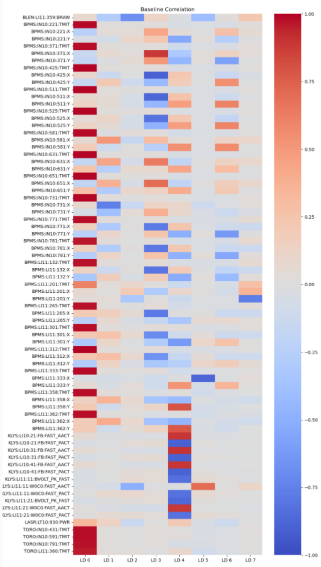
Latent Space Analysis

Latent Space: how it can be used

The learned latent space contains lots of information about the state of the machine:

- We can look at how its axis are correlated to the initial PVs to get a sense of the modes of variation
- We can see how these correlations evolve over time during retraining to get a more direct sense of what drifted

Variation Modes

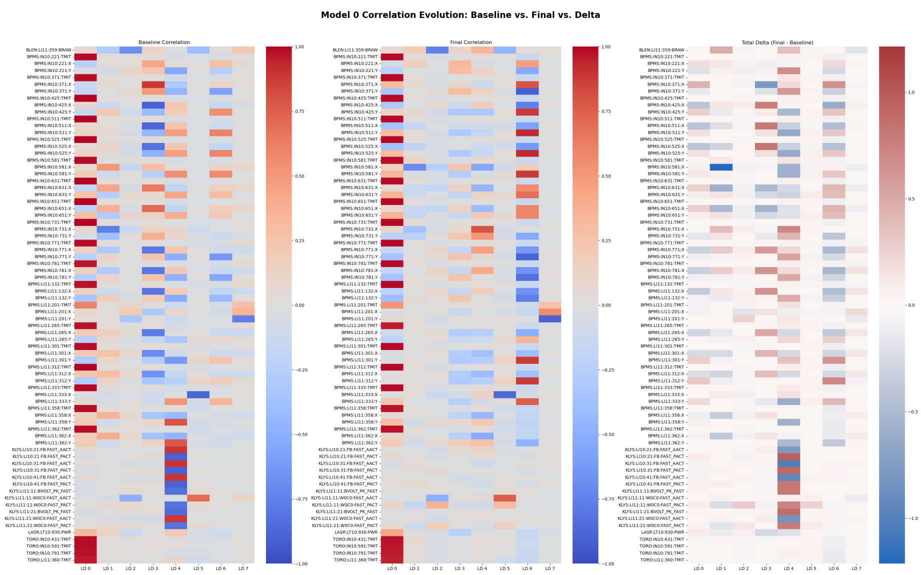


Each VAE learns its own latent space, based on how many specified dimensions are specified by the user.

In the image we can see how the model learned physical meaningful modes, such as correlations between the charge diagnostics.

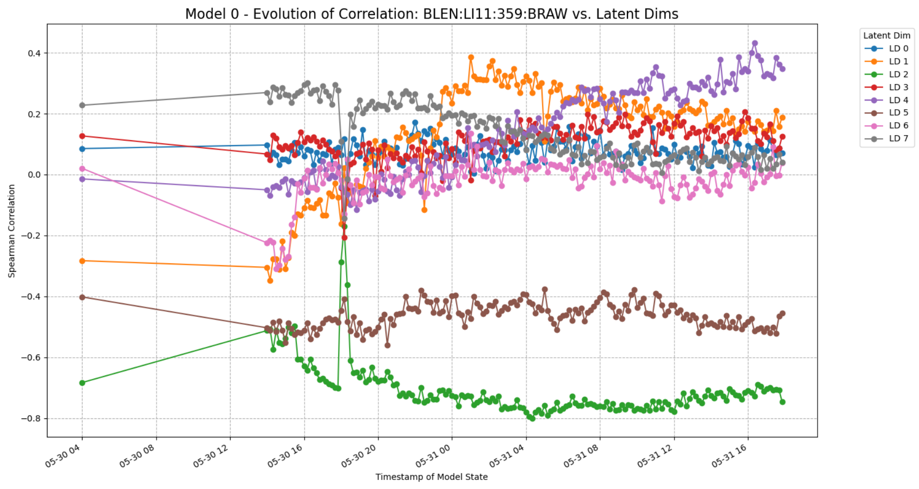
Next, a quick overview of how these correlations evolve over time; an indirect drift detection can be performed by looking at the next plots.

Latent Space Evolution



Model 0 Correlation Evolution: Baseline vs. Final vs. Delta

Latent Space Evolution for a single PV



Conclusion & Outlook

Achievements

- Adaptive VAE framework for FACET-II anomaly detection
- Handles concept drift via hybrid detection + fine-tuning
- Detects real anomalies (RF jumps, jitter, orbit sensitivity)
- Latent space mapped to physical beam/machine parameters

Limitations

- No temporal modeling (shot-to-shot only)
- Manual latent interpretation

Future Work

- Live integration with EPICS, multi-timescale models
- Applications beyond accelerators (industry, energy, aerospace)

Thank you for your attention.