Fault prediction in Wind Turbines via Joint-Embedding Predictive Architecture

David Geddam
Bob Jones University
dgedd236@students.bju.edu

Abstract

Background: In today's generation, wind energy faces significant operation challenges from unpredictable weather conditions, mechanical failures, etc. Traditional wind turbine monitoring relies on large amounts of labeled failure data; this sometimes leads the models to a "memorization gap", where the models fail to identify mechanical drifts. This research addresses the need for a self-supervised world model capable of learning the physics of different wind turbine dynamics before they escalate into catastrophic events [1]. I propose W-JEPA, a Physics-Informed Joint-Embedding Predictive Architecture [1] designed to learn the physics of the turbine operations without labeled data. Unlike standard machine learning models that waste capacity by clustering different features, or deep learning models like variational auto encoders that reconstruct raw sensor noise, W-JEPA operates in a latent space to predict how the turbine's internal behavior changes over time shifts [1, 2].

Methodology: Mainly, our architecture introduces three methods supported by mathematical regularization:

- **Modality Mapping:** To manage the vast high-dimensional input of over 4,755 features, we implement a modular encoder E. Features are grouped into physical subsystems $S \in \{\text{thermal, electrical, dynamics, control}\}$ using regex-based rules. Here, each physical subsystem is projected through a linear layer P_S, such that the global latent state z is formed by aggregating different physical representations:

$$z = \text{Pool}\Big(\text{Concat}(P_{\text{thermal}}(x_{\text{thermal}}), \ldots, P_{\text{control}}(x_{\text{control}}))\Big) \tag{1}$$

- **VICReg Regularization:** To train the model without decoder and prevent model collapse, we utlize **Variance Invariance-Covariance Regularization** [2]. This ensures the latent space preserves the context via a tripartite loss function $\mathscr{L}_{VR}$:
  1. **Invariance ($\mathscr{L}_{inv}$):** This minimizes the mean-square error between student's prediction $\hat{z}_{t+1}$ and the teacher's representation z_{t+1}.
  2. **Variance ($\mathscr{L}_{var}$):** A small hinge loss forces some minimum standard deviation of $\gamma = 1$ for each latent dimension, avoiding the model from collapsing into a single constant vector.
  3. **Covariance ($\mathscr{L}_{cov}$):** This penalizes the off-diagonal elements of the covariance matrix, forcing all the dimensions to be de correlated and to prevent repeating information capture.

- **Physics-Informed Anchor:** The current world model is anchored in physical reality through a secondary objective: actually predicting the **Active Power Output** (P). This is supported by **Cube's Law** ($P \propto v^3$) [4], where v represents wind speed. Here, we apply a high penalty ($\lambda = 15.0$) to this loss to confirm that the latent space prioritizes the energy-conversion dynamics, leading to stability. Final objective function is a weighted sum:

$$\mathscr{L}_{\text{Total}} = \mathscr{L}_{\text{VICReg}} + \lambda \cdot \left\|P_{\text{actual}} - \text{Head}_{\text{power}}(z)\right\|_2^2 \tag{2}$$

Results & Conclusion: Finally, the model was trained on over 32,000 batches of healthy turbine telemetry data from the CARE Dataset [3]. We found out that an evaluation on Asset 50 (turbine with verified hydraulic failure) yielded an **AUROC** (measures how well a model can distinguish between two classes, in our case: "Healthy" vs. "Anomaly") score of 1.00, indicating separation between the normal and anomaly data. Also, W-JEPA provided a 20-day lead time, identifying "surprise" spikes nearly three weeks before baseline models like TimesNet reacted to the failure. This shows that W-JEPA effectively learns a world model of turbine physics, and can be used for predictive maintenance in renewable energy. Limitations include that this was training was conducted for only 20 epochs due to low VRAM and GPU constraints, and thus, future experiements may show higher impact conducting on more epochs, and larger batch sizes to further stabilize VICReg covariance terms and refine the world model's predictions.

Code available on GitHub: https://github.com/dave21-py/Predictive-maintenance-project

Keywords: World Models, JEPA, VICReg.

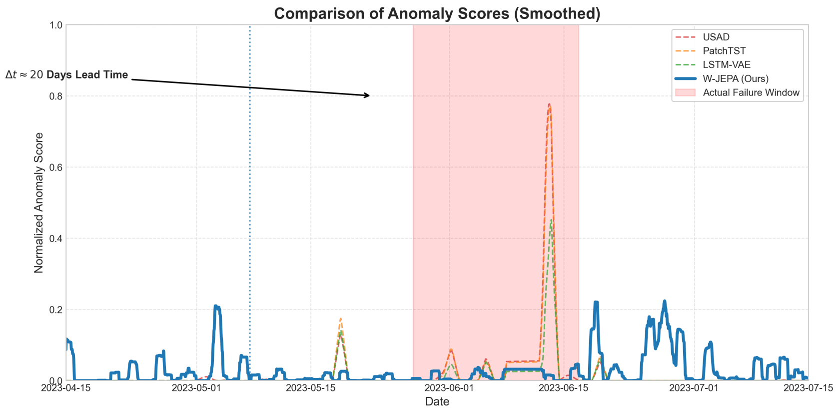

Figure 1: Comparative analysis of latent surprise scores. **May 1st - May 7th:** W-JEPA model (blue) identifies a anomalous mechanical drift on May 7th, and the "surprise" score spiked, later triggering our alarm (dotted line). **May 8th - June 10th:** Once the drift starts, the turbine enters a new physical state, it might be broken or stable in its broken state. Since JEPA is a Predictive Architecture, once it predicts that the turbine is drifting, the surprise stays high relative to the healthy baseline, but does not necessarily spike unless the physics change again. **After June 15th**, the turbine is physically failing, and now even the other baselines such as USAD, PatchTST, LSTM-VAE have their biggest spikes, because the failure became obvious, whereas in contrast to W-JEPA, it already knew about it 20 days earlier.

References

[1] LeCun, Y. (2022). A Path Towards Autonomous Machine Intelligence. *OpenReview*.

[2] Bardes, A., Ponce, J., LeCun, Y. (2021). VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. *arXiv:2105.04906*.

[3] Gück, C., Roelofs, C. M. A., Faulstich, S., (2024). CARE to Compare: A Real-World Benchmark Dataset for Early Fault Detection in Wind Turbine Data. *arXiv:2404.10320*.

[4] Wikipedia. (2026) Square-cube law. *Retrieved January 22, 2026*, from `https://en.wikipedia.org/wiki/Square\%E2\%80\%93cube_law`