Physics-Informed Machine Learning for Robust Industrial Diagnostics: A Systematic Investigation Using Heat Pump Systems

Savvas Eftychis[1], Sławomir Nowaczyk[2] and Sepideh Pashami[3]

[1,2,3] *Center for Applied Intelligent Systems Research, Halmstad University, Halmstad, 30118, Sweden*
savvas.eftychis@hh.se
slawomir.nowaczyk@hh.se
sepideh.pashami@hh.se

1. Introduction & Motivation

Industrial prognostics systems must operate reliably under real-world constraints: limited labeled data, shifting operational conditions, and deployment across heterogeneous units. While Industrial Internet of Things (IIoT) -enabled systems generate vast sensor data, purely data-driven approaches lack the robustness to exploit it effectively, as they tend to overfit to training distributions and fail when conditions change. Physics-Informed Machine Learning (PIML) offers a principled solution by grounding learned models in physical laws, making them more transferable and interpretable.

This thesis investigates whether physical knowledge can provide a robust foundation for industrial diagnostics, using heat pump systems to study generalization across operating conditions and units.

2. Problem Context and Related Work

Out of Distribution (OOD) robustness has received considerable attention in industrial fault diagnosis and condition monitoring, where distribution shift from changing operating conditions and configurations is a well-documented failure mode for deployed data-driven models (Guo et al., 2024). Here, OOD denotes within-unit operating-condition shift while domain generalization (DG) refers to cross-unit shift. DG methods for fault diagnosis have pursued cross-domain invariance through three broad routes: data augmentation across synthetic source domains, domain-invariant representation learning via adversarial or statistical alignment, and learning strategies that regulate optimization across source environments (Zhao, Zio, & Shen, 2024). Performance remains highly task-dependent, suggesting that invariances learned purely from data depend strongly on the diversity of training environments.

Heat pumps provide a representative setting for both OOD and DG. Their operation spans wide ranges of ambient temperature, load, and refrigerant state, while deployments often consist of heterogeneous fleets differing in size, manufacturer, and installation context. Models trained on one regime or unit therefore frequently encounter distribution shift when deployed elsewhere.

PIML offers an alternative source of invariance: physical knowledge. Rather than inferring invariant structure from multiple source domains, PIML embeds known physical relationships directly into the learning process. Physical knowledge can enter as observational bias (inputs, simulation), inductive bias (architecture), or learning bias (training objectives)(Karniadakis et al., 2021).

In prognostics and health management specifically, PIML has been used to improve interpretability, data efficiency, and diagnostic accuracy across a range of applications (Fink, Sharma, et al., 2026; Fink, Nejjar, et al., 2026). Among the available mechanisms, physics-engineered feature construction has been widely adopted, embedding physical laws directly into the feature space (Li, Hu, Liu, Fang, & Kang, 2021). (Niresi, Bissig, Baumann, & Fink, 2024) showed that enriching graph neural network inputs with physics-derived quantities improves performance of virtual sensors in district heating networks. (Wang, Xia, et al., 2025) found that the majority of the most discriminative features for heat pump fault classification were physics-informed. In their work, (Wang, Chen, Guo, Xu, & Chen, 2025) combine thermodynamic constraints with a generative adversarial network to augment sparse data, for the chiller coefficient of performance (COP) prediction, improving the OOD and DG robustness of their prediction.

The first study examined the role of physics-informed features in OOD robustness, in the context of virtual sensing of compressor isentropic efficiency under sensor reduction. A feature space combining raw sensors and thermodynamic transformations was constructed. Neural network models were trained across feasible feature combinations and four sensor removal scenarios, with generalization evaluated using a

PCA-based split that isolates unseen operating regions. Two findings emerged. First, feature composition dominated OOD performance, with variation across feature sets exceeding that attributable to hyperparameter tuning. Second, physics-derived features, particularly entropy, consistently reduced OOD error up to 70% relative to raw-sensor baselines at a modest cost in ID accuracy. These results add to the existing evidence that physics-informed feature engineering improves OOD performance, and further indicate that, within the architecture tested, gains from feature composition can exceed those obtainable through model-side optimization.

The preliminary work establishes the value of physics-informed inputs within a fixed architecture but leaves several questions open. More broadly, while PIML has been widely applied in PHM and DG fault diagnosis has been developed for cross-domain robustness, the two literatures have evolved largely independently: physics priors are rarely evaluated under explicit distribution shift, and DG methods rarely exploit the structural invariances available from physical principles. This thesis addresses these gaps by investigating physics-aligned invariance as a route to OOD and DG robustness in industrial diagnostics.

3. RESEARCH PLAN AND PROPOSED STUDIES

The preliminary work established that features derived from thermodynamics (entropy in particular) reduced OOD error substantially relative to raw-sensor baselines. The thesis tests whether the same principle extends from inputs to learned representations, and from single-unit to cross-unit settings, through three contributions. Studies 2 and 3 examine physics-aligned latent spaces for OOD and DG robustness respectively.

3.1. Physics-Structured Latent Spaces for Robust Industrial Diagnostics

This study will investigate whether physical knowledge can be incorporated at the level of the learned representation. In the proposed approach, an encoder will be trained so that the pairwise distances between samples in its latent space match the differences in a thermodynamic key performance indicator (KPI). COP will be used as the alignment target, computed at training time from the available sensor data, as its variation is broadly indicative of fault states (Cui & Wang, 2005). The alignment will follow the supervised contrastive principle for continuous targets introduced by Taghiyarrenani et al. (Taghiyarrenani, Nowaczyk, Pashami, & Bouguelia, 2023): samples with similar COP values are placed close together in latent space, while samples with dissimilar COP values are placed proportionally further apart.

Training will draw on multiple data sources that share the underlying physics. Beyond the single unit, the alignment will be informed by data from additional units with the same design and by calibrated simulations whose parameters are matched to real-data operating coefficients. Evaluation remains single-unit, with the auxiliary sources populating sparsely-sampled regions of the KPI range without entering the held-out distribution.

The encoder is paired with a task head and trained jointly under a combined loss that includes both the contrastive alignment term and the task loss, so that the latent space is shaped by both the COP geometry and the task. This architecture is instantiated separately for each downstream task examined. The first case is KPI regression, the most directly evaluable on real data and the carrier of the primary OOD claim. Additional tasks include residual-based anomaly detection and fault classification using simulation labels. COP changes alone are limited in diagnostic content, typically scalar increases or decreases, but the alignment constrains pairwise distances rather than coordinates, so the latent space inherits COP's physical interpretability in its metric while remaining free to use orthogonal dimensions. Whether those dimensions can carry information sufficient for the downstream task is one of the questions this study will address.

3.2. Physics-Aligned Domain Generalization Across Heat Pump Units

The final study will extend the latent-space alignment approach to DG. Unlike Study 2, units are held out during training and used as transfer targets. The central hypothesis is that normalised thermodynamic KPIs provide transferable structure across units.

Study 3 reuses the architecture and tasks from Study 2. Cross-unit evaluation will combine real and simulated units spanning a range of design characteristics. Both will be used as training sources and held-out targets, enabling evaluation under both interpolation and extrapolation. As an extension, symbolic regression on the cross-unit aligned latent space will produce candidate symbolic expressions relating sensor inputs to latent coordinates, providing domain experts with interpretable equations of the learned structure.

3.3. Evaluation Protocol

The two studies share a common evaluation framework, with study-specific instantiations.

Data: Real operational data from partner heat pumps and calibrated simulations containing healthy and faulty samples.

Splits: Study 2 reuses the PCA-based split methodology from the preliminary work to isolate unseen operating regions of the focal unit. Alternative split strategies will also be explored. Study 3 evaluates on held-out units, selected to span similar and dissimilar designs relative to the training set, so that performance can be assessed under both modest and substantial design shift.

Tasks: KPI regression, residual-based anomaly detection, and fault classification. A common encoder with task-specific heads will be used to assess whether the learned representation transfers across multiple downstream PHM tasks. Standard metrics will be used for each task.

Baselines: The primary comparison is against empirical risk minimisation (ERM): the same regression head trained without the pairwise-distance alignment, on raw and physics-informed inputs equivalent to the configurations evaluated in the preliminary work. The preliminary's best feature-engineering configuration provides a second reference point.

3.4. Timeline and Milestones

Year 1. Preliminary work; PHME 2026 paper 1 (completed).
Year 2. Simulation infrastructure, Study 2 alignment on real data, Study 2 augmented with simulations, paper 2.
Year 2-3. Multi-unit data and design-varied simulations prepared, Study 3 cross-unit alignment, paper 3.
Year 4. Symbolic regression extension, paper 4.

3.5. Expected Contributions

A central contribution of this thesis is the development and evaluation of physics-aligned latent representations for industrial machine learning. Rather than relying solely on invariances discovered from data, the proposed approach investigates how known physics can be used to shape embedding spaces, creating representations that remain meaningful under distribution shift and across heterogeneous units. Through the progression from physics-informed features (Study 1) to physics-aligned latent spaces (Studies 2 and 3), the thesis aims to establish a principled understanding of how physical knowledge can be used to improve robustness in industrial diagnostics. A complementary contribution is improved explainability through symbolic regression, which will be used to identify physically meaningful relationships within the learned representation

REFERENCES

Cui, J., & Wang, S. (2005). A model-based online fault detection and diagnosis strategy for centrifugal chiller systems. *International Journal of Thermal Science*, *10*(44), 986-999.

Fink, O., Nejjar, I., Sharma, V., Niresi, K. F., Sun, H., & Dong, . . K. Y., H. (2026). From physics to machine learning and back: Part ii-learning with inductive biases in prognostics and health management. *Reliability Engineering System Safety*, 112376.

Fink, O., Sharma, V., Nejjar, I., Von Krannichfeldt, L., Garmaev, S., & Zhang, . . S. K., Z. (2026). From physics to machine learning and back: Part i-learning with inductive biases in prognostics and health management. *Reliability Engineering System Safety*, 112213.

Guo, Y., Wang, N., Shao, S., Huang, C., Zhang, Z., Li, X., & Wang, Y. (2024). A review on hybrid physics and data-driven modeling methods applied in air source heat pump systems for energy efficiency improvement. *Renewable and Sustainable Energy Reviews*, *204*, 114804.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*, 422-440.

Li, G., Hu, Y., Liu, J., Fang, X., & Kang, J. (2021). Review on fault detection and diagnosis feature engineering in building heating, ventilation, air conditioning and refrigeration systems. *IEEE Access*, *9*, 2153-2187.

Niresi, K. F., Bissig, H., Baumann, H., & Fink, O. (2024). Physics-enhanced graph neural networks for soft sensing in industrial internet of things. *IEEE Internet of Things Journal*, *11*(21), 34978-34990.

Taghiyarrenani, Z., Nowaczyk, S., Pashami, S., & Bouguelia, M. R. (2023). Multi-domain adaptation for regression under conditional distribution shift. *Expert systems with applications*, *224*, 119907.

Wang, Z., Chen, J., Guo, K., Xu, B., & Chen, Z. (2025). Study on data augmentation with physics-informed generative adversarial networks and the extrapolation performance of cop prediction for chillers. *Energy Conversion and Management*, *346*, 120418.

Wang, Z., Xia, P., Guo, J., Zhou, S., Wang, L., Wang, Y., & Zhang, C. (2025). Efficient feature selection for enhanced chiller fault diagnosis: A multi-source ranking information-driven ensemble approach. *Building Simulation*, *18*, 141-159.

Zhao, C., Zio, E., & Shen, W. (2024). Domain generalization for cross-domain fault diagnosis: an application-oriented perspective and a benchmark study. *Reliability Engineering System Safety*, *245*, 109964.