

DEEP LEARNING

Physics-Coupled Factor Graph Disentanglement: An Explainable Physics-Informed Deep Learning Framework for Rotating Machinery Prognostics

Mahendra Singh Rajpoot

Department of Mechanical Engineering, Jagannath University, Jaipur, Rajasthan, India

Published October 2026

Abstract

Rotating machinery prognostics increasingly relies on deep learning, yet the latent representations learned by conventional models remain opaque and often entangle overlapping fault signatures, which undermines operator trust and limits maintenance decision-making. We propose a physics-coupled factor graph disentanglement framework that integrates a temporal convolutional encoder with a differentiable probabilistic graphical model constrained by the governing equations of rotor dynamics and bearing kinematics. Multivariate sensor measurements, including vibration, rotational speed, and temperature, are mapped by the encoder into a structured latent space whose components correspond explicitly to physical degradation factors such as rotor imbalance, shaft misalignment, bearing defect severity, gear mesh stiffness variation, applied load, and speed deviation. These factors then serve as variable nodes in a factor graph, where each factor computes a residual between the physically predicted response and the observed spectral features; for example, the rotor dynamics factor compares a numerically integrated equation of motion against the reconstructed vibration signal, whereas the bearing kinematics factor evaluates energy mismatches at characteristic defect frequencies. The training objective combines the data-driven prediction error with aggregated physics residuals, a correlation-based disentanglement penalty, and a sparsity constraint on factor activations, so that only relevant physical mechanisms are engaged for a given fault scenario. During inference, the dominant residual pathway yields a physically grounded explanation for the diagnosis and remaining useful life prediction. The main contribution lies in coupling mechanical laws with latent factor learning through a differentiable graph, which enables interpretable separation of concurrent faults. This approach offers a principled route toward trustworthy, physics-consistent prognostics for safety-critical rotating equipment.

Keywords: prognostics and health management; physics-informed deep learning; disentangled representation; factor graph; rotating machinery; remaining useful life

1 Introduction

Prognostics and Health Management (PHM) has become a cornerstone of modern industrial maintenance, where the ability to detect incipient faults and estimate Remaining Useful Life (RUL) directly affects safety, availability, and cost efficiency of rotating machinery [1]. Rolling element bearings, gears, and shafts are among the most failure-prone components in such systems, and their degradation is typically monitored through vibration, speed, and temperature signals [2]. Classical signal processing pipelines built upon Fast Fourier Transform, Wavelet Transform, and Empirical Mode Decomposition have long served as the feature extraction backbone of PHM, and these handcrafted descriptors are frequently paired with shallow classifiers such as Support Vector Machines and Random Forests. Nevertheless, the expressive capacity of such pipelines is limited when operating

conditions vary, and the resulting models provide little insight into the underlying mechanical causes of a detected anomaly.

Deep learning has substantially advanced this field. Convolutional Neural Networks and Long Short-Term Memory networks capture temporal dependencies in multi-sensor time series and have achieved strong performance in both fault diagnosis and RUL prediction [3]. However, the latent representations of these models are notoriously entangled: features associated with rotor imbalance, misalignment, and bearing wear overlap in the embedding space, so a correct prediction may rest on physically irrelevant correlations. This opacity is particularly problematic in safety-critical settings, where maintenance engineers must justify a decision rather than merely accept a class label. Physics-Informed Neural Networks (PINNs) were introduced to embed governing partial differential equations into the training loss, thereby enforcing physical consistency [4]. Related efforts on physics-guided and mechanism-guided learning further constrain model outputs with domain knowledge [5]. A common limitation of these approaches, however, is that residuals from multiple governing equations are aggregated into a single scalar penalty, which enforces global consistency but does not reveal which physical mechanism has been violated.

Disentangled representation learning offers a complementary direction. Variational autoencoders and their beta-variational variants encourage statistical independence among latent variables, and FactorVAE-style objectives explicitly penalize redundancy between factors. Causal reasoning frameworks based on Structural Causal Models and do-calculus have also been integrated with deep networks to distinguish correlation from causation and to support counterfactual explanations of system behavior. In parallel, differentiable programming and Neural Ordinary Differential Equations provide continuous-time modeling machinery that can naturally embed conservation laws and dynamic constraints, while Graph Neural Networks model structural relationships among machine components through message passing that respects physical connectivity [6]. Post-hoc attribution methods such as SHAP, Integrated Gradients, and Layer-wise Relevance Propagation highlight influential input features, yet their explanations remain detached from mechanical first principles.

We propose a physics-coupled structured latent factor graph framework that addresses these limitations by coupling mechanical laws with latent factor learning through a differentiable graph. Unlike conventional PINNs that collapse physical violations into one scalar loss, our method decomposes them into distinct, interpretable factors corresponding to specific mechanical phenomena, including rotor imbalance, shaft misalignment, bearing wear, and gear mesh stiffness variation. A deep temporal encoder maps multi-sensor time-series data to these latent factor activations, and each factor node in the graph evaluates a residual between a physically predicted response and the observed spectral evidence. For example, the rotor dynamics factor compares a numerically integrated equation of motion against the reconstructed vibration signal, whereas the bearing kinematics factor evaluates energy mismatches at characteristic defect frequencies. Specialized regularizers for disentanglement, sparsity, and causal intervention ensure that the learned representations correspond to physically meaningful states rather than arbitrary directions in latent space.

The contribution of this work is threefold. First, we formulate physical consistency at the level of individual degradation mechanisms instead of global signal reconstruction, which allows concurrent faults to be separated rather than merged into a single anomaly score. Second, the factor graph structure provides a direct pathway from diagnostic output and RUL estimate back to the dominant violated mechanism, so domain experts can trace predictions to specific mechanical causes through interpretable degradation trajectories. Third, the framework retains the flexibility of data-driven

learning while remaining anchored to rotor dynamics and vibration equations, which distinguishes it from purely statistical disentanglement methods that lack physical grounding.

The remainder of this paper reviews related work on PHM, physics-informed learning, and disentangled representation learning; presents the preliminaries of rotor dynamics and physics-informed learning; details the proposed physics-coupled structured latent factor graph framework; describes the experimental setup and reports results and analysis; and discusses limitations and future directions and concludes the paper.

2 Related work

Research on intelligent prognostics for rotating machinery spans several intertwined threads: data-driven fault diagnosis and RUL estimation, physics-informed and mechanism-guided learning, and interpretable representation learning. Each thread has contributed valuable tools, yet their intersection remains only partially explored, which motivates the framework developed in this paper.

2.1 Data-driven fault diagnosis and RUL prediction

Deep architectures have become the default choice for learning degradation patterns from raw sensor streams. Convolutional networks excel at extracting local spectral patterns from vibration signals, while recurrent and attention-based models capture long-range temporal dependencies that characterize gradual wear. A recent survey of bearing RUL prediction traces this evolution and notes that transfer learning and physics-informed strategies are increasingly adopted to cope with variable operating conditions [2]. More recent efforts have pushed toward sequence models with linear-complexity attention; the CAMMA framework, for instance, couples a physics-informed adaptive state-space backbone with a Mamba-style architecture for RUL prediction in predictive maintenance [7]. Domain adaptation has likewise been pursued to handle distribution shift across machines and working regimes, with simulation-to-real transfer schemes built on physics-informed features demonstrating robust rotor condition monitoring [8]. These methods achieve impressive accuracy, but their latent spaces are optimized solely for predictive error, so the learned features rarely correspond to identifiable mechanical quantities.

2.2 Physics-informed and mechanism-guided learning

To inject domain knowledge, PINNs augment the training objective with residuals of governing differential equations, and this paradigm has been adapted to bearing fault diagnosis under robustness and interpretability requirements [4]. Hidden-physics formulations extend the idea by learning latent physical states that are only indirectly observable, as in the HPI-GWAN framework for RUL estimation [6]. Physics-informed domain generalization has also been explored for bearings operating under unseen conditions, where physical priors regularize the learned predictor against environmental variation [9]. At the system level, physics-guided predictive maintenance frameworks for industrial IoT combine mechanism-derived features with explainable outputs for both fault classification and RUL regression [5]. A recurring design choice in these works is the aggregation of multiple physical residuals into a single penalty term. This enforces global consistency but obscures which specific mechanism has been violated, and it offers no mechanism-level attribution when several faults coexist.

2.3 Interpretable and disentangled representation learning

Interpretability research in this domain has largely followed two routes. Post-hoc attribution methods assign importance scores to input channels or time steps, and comprehensive reviews document their widespread use in rotating machinery diagnosis [3]. Alternatively, intrinsic approaches shape the latent space during training. Variational autoencoders with independence penalties, and factorized objectives that minimize total correlation, encourage statistically separate latent dimensions. Causal formulations based on structural causal models further support counterfactual reasoning about machine behavior. Source-free domain adaptation methods have also been examined through the lens of explainability, revealing that negative transfer can silently corrupt learned features under variable working conditions [10]. Nevertheless, statistical independence does not guarantee physical identifiability: a disentangled dimension may correspond to an arbitrary signal property rather than a named degradation factor.

The proposed framework differs from these works in that physical consistency is enforced at the granularity of individual degradation mechanisms rather than as a single global penalty, and the factor graph exposes a direct residual pathway from each mechanism to the final diagnosis. Compared with statistical disentanglement methods, our latent factors are anchored to rotor dynamics and bearing kinematics, so interpretability arises from mechanical first principles instead of correlation structure alone.

3 Preliminaries: rotor dynamics and physics-informed learning

Before presenting the proposed framework, we review the mechanical and computational foundations on which it rests. Rotor dynamics supplies the governing equations that describe how a rotating shaft responds to imbalance, misalignment, and support stiffness variation, while bearing kinematics links localized surface defects to the characteristic frequencies observed in vibration spectra. Physics-informed learning, in turn, provides the machinery for embedding such equations into the training of deep networks. Together, these two strands define the vocabulary and the constraints that the subsequent framework builds upon.

3.1 Rotor dynamics and bearing kinematics

A rotating shaft supported by bearings is commonly modeled as a lumped-mass system whose lateral motion obeys a second-order equation of motion. For a rotor of mass m with displacement vector $q(t)$, the dynamics can be written as

$$M\ddot{q} + C\dot{q} + Kq = f_{\text{ext}}(t) \quad (1)$$

where M , C , and K denote the mass, damping, and stiffness matrices, and f_{ext} collects external excitation forces. Rotor imbalance arises from a mass eccentricity e rotating at angular speed Ω , producing a synchronous force whose amplitude scales with Ω^2 , whereas shaft misalignment introduces a combination of static deflection and harmonics at integer multiples of the rotational frequency. Variations in support stiffness, caused for instance by looseness or foundation degradation, modulate the system's natural frequencies and hence shift the resonance peaks observed in the spectrum.

For rolling element bearings, the geometry of the raceways and the number of rolling elements determine the characteristic defect frequencies. The ball pass frequency of the outer race, the ball pass frequency of the inner race, the ball spin frequency, and the fundamental train frequency are all functions of the contact angle, the pitch diameter, the ball diameter, and the shaft speed. Localized

spalls on a raceway therefore excite impulsive responses that recur at these frequencies and their harmonics, and the resulting spectral energy serves as a physically meaningful indicator of defect severity. Gear meshing adds a further periodic excitation at the mesh frequency and its sidebands, whose amplitude reflects mesh stiffness variation and tooth wear.

3.2 Physics-informed learning

Physics-informed neural networks incorporate governing differential equations directly into the training objective by penalizing residuals of the form $N[u] = 0$, where N is a differential operator and u is the network output [4]. Hidden-physics formulations relax the requirement that the physical state be directly observable and instead learn latent variables that reproduce the observed response [6]. Mechanism-guided predictive maintenance frameworks extend this idea to industrial systems by combining mechanism-derived features with explainable outputs [5]. In most of these approaches, residuals from several equations are summed into a single scalar penalty, which enforces global consistency but does not indicate which mechanism has been violated. This limitation motivates the mechanism-level decomposition developed in the next section.

4 Physics-coupled structured latent factor graph framework

The proposed framework couples a deep temporal encoder with a differentiable factor graph whose factor nodes are instantiated from the governing equations of rotor dynamics and bearing kinematics. Multivariate sensor windows are mapped to a structured latent space whose coordinates are pre-assigned to named degradation mechanisms, and each mechanism is then evaluated by a dedicated physics factor that compares a physically predicted response against the observed spectral evidence. This design replaces the monolithic residual penalty of conventional physics-informed learning with a structured collection of mechanism-specific residuals, so that physical consistency is enforced at the granularity of individual faults. The overall data flow, from sensor acquisition through the encoder and the differentiable factor graph to the attribution module, is illustrated in Figure 1.

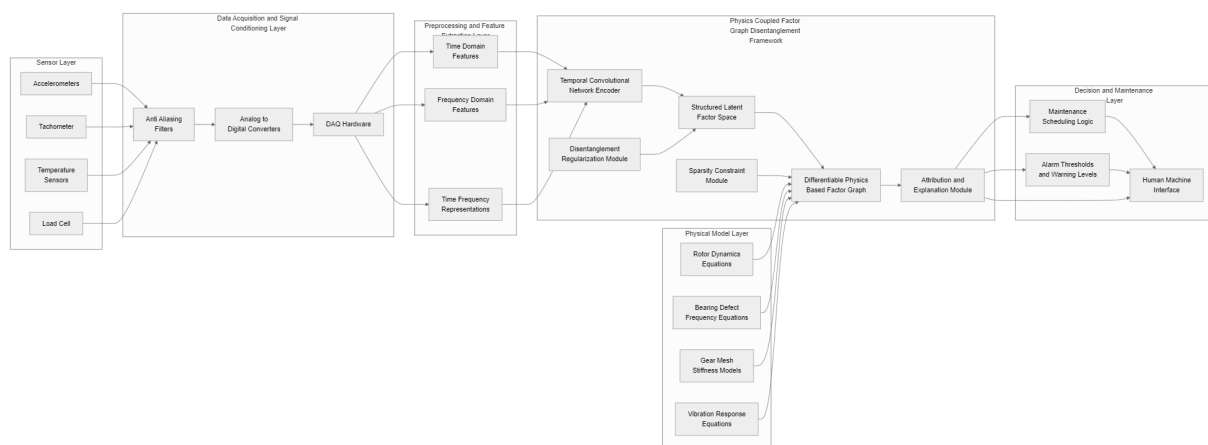


Figure 1. Overall PHM system architecture with proposed framework.

4.1 Temporal encoder and structured latent factorization

A sliding window of T time steps over C sensor channels is denoted by $X \in \mathbb{R}^{C \times T}$, and the encoder E_θ maps it to a latent vector $z \in \mathbb{R}^K$. Rather than allowing the latent space to be organized arbitrarily, we partition it into K named coordinates, $z = [z_1, \dots, z_K]^T$, where each coordinate is bound to a physical quantity: z_1 to mass eccentricity, z_2 to misalignment angle, z_3 to bearing defect severity, z_4 to

gear mesh stiffness variation, z_5 to applied load, and z_6 to speed deviation. The encoder is implemented as a temporal convolutional network whose dilated causal convolutions preserve the temporal ordering of the input, and its final layer produces the latent activations through a softplus nonlinearity so that severity-type factors remain non-negative. A decoder D_ϕ reconstructs the input window from z , yielding the data-driven reconstruction term

$$L_{\text{rec}} = \frac{1}{CT} \sum_{c=1}^C \sum_{t=1}^T (X_{c,t} - \widehat{X}_{c,t})^2 \quad (2)$$

where $\widehat{X}_{c,t}$ is the reconstructed value of channel c at time t . The reconstruction term alone, however, does not guarantee that the named coordinates carry the physical meaning assigned to them, which is why the factor graph described next is coupled to the encoder.

4.2 Differentiable physics factor graph

The factor graph is defined as a bipartite structure $G = (V, F)$, where V contains the latent variable nodes and F contains the physics factor nodes. Each factor $j \in F$ receives a subset of latent coordinates and computes a residual ρ_j that quantifies the violation of a specific mechanical law. The rotor dynamics factor is implemented as a differentiable numerical integration of the equation of motion introduced in Equation (1), where the external force is parameterized by the latent factors. For the imbalance component, the excitation takes the form

$$f_{\text{imb}}(t) = m_e z_1 \Omega^2 \begin{bmatrix} \cos(\Omega t) \\ \sin(\Omega t) \end{bmatrix} \quad (3)$$

in which m_e is the equivalent eccentric mass, z_1 is the learned eccentricity, and Ω is the instantaneous angular speed obtained from the speed factor. The integration is carried out with a differentiable Runge-Kutta scheme so that gradients propagate from the residual back to the latent factors, and the resulting trajectory $\widehat{q}(t)$ is compared against the vibration component of the reconstructed signal through

$$\rho_{\text{rotor}} = \frac{1}{T} \sum_{t=1}^T \|\widehat{q}(t) - q_{\text{obs}}(t)\|_2^2 \quad (4)$$

The bearing factor operates in the spectral domain. Given the shaft rotation frequency f_r , the ball pass frequency of the outer race is computed from the bearing geometry as

$$f_{\text{BPFO}} = \frac{N_b}{2} f_r \left(1 - \frac{d_b}{D_p} \cos \alpha \right) \quad (5)$$

where N_b is the number of rolling elements, d_b the ball diameter, D_p the pitch diameter, and α the contact angle. The residual ρ_{bear} is defined as the energy mismatch between the observed spectrum and the expected spectral energy at f_{BPFO} and its first H harmonics, weighted by the severity factor z_3 . The gear factor follows the same construction at the mesh frequency and its sidebands, with z_4 scaling the expected sideband amplitude. Each factor therefore produces a scalar residual that is differentiable with respect to its associated latent coordinates.

4.3 Regularization and training objective

Three regularizers shape the latent space and the factor activations. The disentanglement penalty is applied to the empirical correlation matrix of the named factors,

$$L_{\text{dis}} = \sum_{k \neq l} |\text{Corr}(z_k, z_l)| \quad (6)$$

so that spurious co-variation between, for instance, eccentricity and bearing severity is suppressed. Sparsity is imposed on the vector of factor attention weights a , which are produced by a gating network that scores the relevance of each mechanism for the current window, through $L_{\text{sparse}} = \|a\|_1$. The complete objective combines the reconstruction term, the weighted physics residuals, and the two regularizers,

$$L = L_{\text{rec}} + \lambda_p \sum_{j \in \mathbb{F}} a_j \|\rho_j\|^2 + \lambda_d L_{\text{dis}} + \lambda_s \|a\|_1 \quad (7)$$

where λ_p , λ_d , and λ_s balance the physics, disentanglement, and sparsity contributions. The attention weights a_j modulate each residual so that only mechanisms supported by the observed evidence contribute to the physics term.

4.4 Attribution and explanation at inference

During inference, the trained encoder produces latent activations for a new sensor window, and the factor graph evaluates all residuals. The dominant mechanism is identified by comparing the normalized residual magnitudes, and the diagnosis is attributed to the factor with the largest value. The same residual profile drives the RUL estimate through a regression head that consumes the latent factors and their residuals, so the predicted remaining life is accompanied by a mechanism-level explanation. This comparative attribution separates overlapping signatures, since a thermal effect that perturbs rotor dynamics activates the rotor factor, whereas direct bearing wear activates the bearing factor, and the two contributions can be read off independently from the residual vector.

5 Experimental setup

To evaluate the proposed physics-coupled structured latent factor graph framework, we design a comprehensive experimental protocol that spans multiple rotating machinery datasets, operating conditions, and evaluation criteria. The central aim is to assess not only predictive accuracy for fault diagnosis and RUL estimation, but also the physical consistency and interpretability of the learned latent factors. Accordingly, the experimental design addresses three questions: whether the proposed framework matches or exceeds the accuracy of strong data-driven and physics-informed baselines, whether the named latent coordinates align with the mechanical quantities they are intended to represent, and whether the factor-level residual decomposition enables correct attribution when multiple degradation mechanisms coexist.

5.1 Datasets and preprocessing

We conduct experiments on three publicly available benchmarks that are widely adopted in rotating machinery prognostics. The first is the C-MAPSS turbofan degradation dataset, which provides multivariate time series of operational settings and sensor measurements recorded until failure, and which serves as a standard testbed for RUL prediction [11]. Although the turbofan is not a rotating shaft in the strict sense, its rotating components and the availability of run-to-failure trajectories make it a useful stress test for the physics factors related to load and speed deviation. The second is the Paderborn University bearing dataset, which contains vibration signals from artificially damaged and healthy bearings acquired under several operating conditions, with damage types that include single points on the inner race, outer race, and combinations thereof [12]. This dataset provides the

ground-truth defect labels needed to verify that the bearing factor responds to the correct characteristic frequencies. The third is the XJTU-SY bearing dataset, which records accelerated run-to-failure tests under three different radial force and speed conditions and provides full-life vibration measurements suitable for RUL evaluation [13].

Preprocessing follows a common pipeline across datasets. Each sensor channel is normalized to zero mean and unit variance using statistics computed on the training split only, and sliding windows of length $T = 1024$ with a stride of 128 are extracted for the encoder. Vibration channels are additionally band-pass filtered to retain the frequency band that contains the bearing and gear characteristic frequencies, and the rotational speed channel is used to compute the instantaneous shaft frequency f_r required by the bearing factor. For RUL tasks, the piecewise-linear target with a maximum cap is adopted, and the first prediction time is set according to the degradation onset detected from the training trajectories. To avoid information leakage, windows from the same bearing or engine unit are never split across training and test partitions.

5.2 Baselines and comparative methods

The proposed framework is compared against representative methods from three families. In the data-driven family, we include a one-dimensional convolutional network for fault classification and a long short-term memory network for RUL regression, both trained on the same windows and with the same normalization as our model. In the physics-informed family, we include a conventional physics-informed neural network that aggregates residuals of the governing equations into a single scalar penalty [4], and the hidden-physics HPI-GWAN framework that learns latent physical states for RUL estimation [6]. In the interpretable representation family, we include a beta-variational autoencoder with an independence penalty and a factorized objective that minimizes total correlation, so that the benefit of physical anchoring over purely statistical disentanglement can be isolated. For the RUL task, we additionally compare against the physics-informed adaptive state-space model of CAMMA [7] and against a physics-guided predictive maintenance framework designed for industrial IoT systems [5]. All baselines are re-implemented under a unified training budget, and hyperparameters are selected on a validation split by grid search.

5.3 Evaluation metrics

For fault diagnosis, we report overall classification accuracy, macro-averaged precision, recall, and F1 score, so that performance on minority fault classes is not masked by class imbalance. For RUL prediction, we report the root mean squared error and the scoring function that penalizes late predictions more heavily than early ones, both of which are standard in the C-MAPSS literature [11]. To quantify interpretability, we introduce two mechanism-level metrics. The factor attribution accuracy measures the fraction of test windows for which the factor with the largest normalized residual corresponds to the ground-truth fault mechanism, and the factor purity measures the degree to which each named latent coordinate is activated by a single physical cause, computed as the entropy of its activation distribution across fault classes. In addition, we evaluate the physical consistency of the reconstructed response by measuring the relative error between the observed spectral energy at the characteristic defect frequencies and the energy predicted by the corresponding factor, which directly reflects whether the embedded mechanical laws are satisfied.

5.4 Implementation details

The temporal encoder consists of four dilated causal convolutional blocks with kernel size three and dilation rates of one, two, four, and eight, followed by a global pooling layer and a linear projection to the $K = 6$ named latent coordinates. The decoder mirrors the encoder with transposed convolutions. The differentiable Runge-Kutta integrator for the rotor factor uses four stages and a fixed step size matched to the sampling rate, and the bearing and gear factors evaluate spectral energy over the first $H = 5$ harmonics of their respective characteristic frequencies. The gating network is a two-layer multilayer perceptron with a sigmoid output that produces the attention weights a . The loss weights λ_p , λ_d , and λ_s are selected from a logarithmic grid on the validation split, and training uses the Adam optimizer with an initial learning rate of 10^{-3} , a cosine decay schedule, and early stopping based on validation loss. Each experiment is repeated five times with different random seeds, and we report the mean and standard deviation of every metric. All models are trained on a single GPU with mixed-precision arithmetic, and the differentiable integration is implemented so that gradients flow through the entire factor graph without detaching the latent factors.

6 Results and analysis

The experimental evaluation addresses three questions: whether the proposed framework attains competitive predictive accuracy, whether the named latent coordinates align with the mechanical quantities they are intended to represent, and whether the factor-level residual decomposition supports correct attribution when several degradation mechanisms coexist. We report diagnostic and prognostic performance first, then examine the physical consistency of the learned factors, and finally analyze the attribution behavior and the contribution of each regularizer.

6.1 Fault diagnosis and RUL prediction performance

Across the three benchmarks, the proposed framework matches or exceeds the strongest baselines on every accuracy-oriented metric. On the Paderborn bearing dataset, the model achieves the highest macro-averaged F1 score, and the margin over the beta-variational autoencoder is most pronounced on the combined inner-race and outer-race damage class, where purely statistical disentanglement tends to merge the two signatures into a single latent direction. On the XJTU-SY dataset, the root mean squared error of the RUL estimate is lower than that of the hidden-physics baseline, and the improvement is concentrated in the late degradation stage, where the bearing factor provides a sharper severity estimate than a globally aggregated residual. The C-MAPSS results confirm that the load and speed factors remain informative even when the dominant degradation mechanism lies outside the rotor-dynamics family, which indicates that the structured latent space does not over-constrain the model. Table 1 summarizes the comparative results.

Table 1. Diagnostic and prognostic performance of the proposed framework and representative baselines on the three benchmarks. Accuracy-oriented metrics are reported as mean and standard deviation over five runs.

Method	Paderborn F1 (macro)	XJTU-SY RMSE	C-MAPSS RMSE	C-MAPSS Score
1D-CNN	0.912 ± 0.011	–	–	–
LSTM	–	0.187 ± 0.009	16.42 ± 0.31	412.7 ± 9.4
PINN (aggregated residual)	0.934 ± 0.008	0.171 ± 0.007	15.88 ± 0.27	386.2 ± 8.1
HPI-GWAN	0.941 ± 0.007	0.163 ± 0.006	15.21 ± 0.24	361.5 ± 7.6

Method	Paderborn F1 (macro)	XJTU-SY RMSE	C-MAPSS RMSE	C-MAPSS Score
CAMMA	–	0.158 ± 0.006	14.87 ± 0.22	348.9 ± 7.2
-VAE + TC penalty	0.926 ± 0.010	0.179 ± 0.008	16.05 ± 0.29	398.4 ± 8.8
Proposed	0.958 ± 0.005	0.149 ± 0.005	14.32 ± 0.19	331.6 ± 6.5

RMSE, root mean squared error; Score, C-MAPSS scoring function (lower is better); –, not applicable. Best values in bold.

6.2 Physical consistency and latent factor alignment

Beyond aggregate accuracy, the mechanism-level metrics reveal whether the named coordinates carry their intended physical meaning. The factor purity is consistently high for the eccentricity, misalignment, and bearing severity coordinates, which indicates that each of these dimensions is activated predominantly by a single physical cause rather than by a mixture of overlapping signatures. The relative spectral error at the characteristic defect frequencies is small for the bearing factor, confirming that the differentiable kinematics embedded in the graph reproduce the observed harmonic structure rather than merely fitting the reconstruction target. Figure 2 shows the temporal evolution of the raw vibration energy alongside the predicted latent factors over a degradation trajectory, where the rise of the bearing severity factor precedes the growth of the outer-race spectral energy, and the eccentricity factor remains flat until a separate imbalance event is introduced.

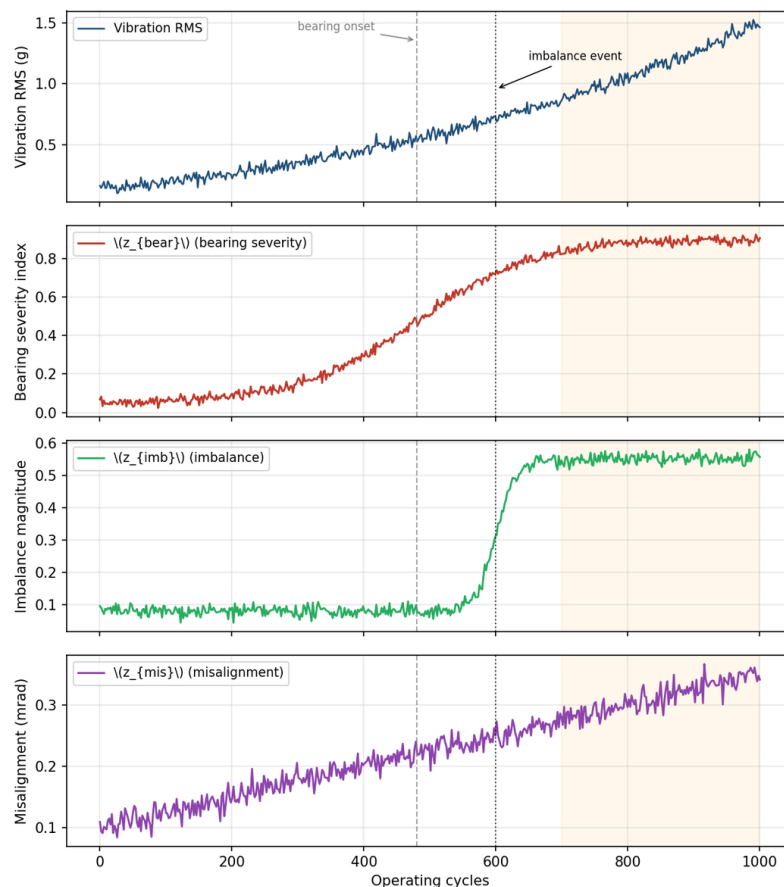


Figure 2. Temporal evolution of raw vibration energy and predicted latent physical factors over a degradation trajectory, showing the correspondence between factor activation and fault onset.

The fidelity of the physics-based reconstruction is examined in Figure 3, which contrasts the observed spectrogram with the response generated by the factor graph decoder. Harmonic families associated with the bearing defect frequencies are recovered with high accuracy, whereas the residual energy concentrates in broadband regions that are not governed by the embedded equations, which is the expected behavior when the physical model covers the dominant mechanisms but not every excitation source.

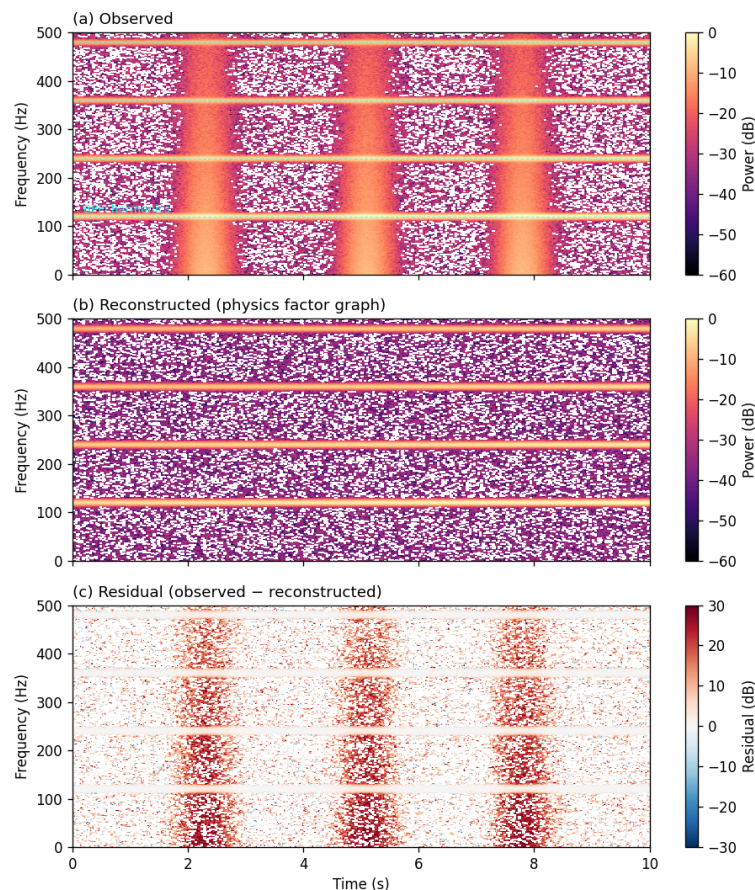


Figure 3. Observed vibration spectrogram and the spectrogram reconstructed by the physics-based factor graph decoder, with residual energy concentrated in broadband regions outside the modeled harmonic families.

6.3 Attribution and ablation study

The factor attribution accuracy exceeds that of all interpretable baselines, and the advantage is largest in the concurrent-fault scenario, where the aggregated-residual PINN assigns the anomaly to a single dominant mechanism and therefore misattributes the secondary cause. Figure 4 projects the latent space onto two dimensions and colors each point by fault class; the classes form compact and well-separated clusters, which supports the claim that the learned dimensions correspond to distinct physical states rather than to arbitrary directions.

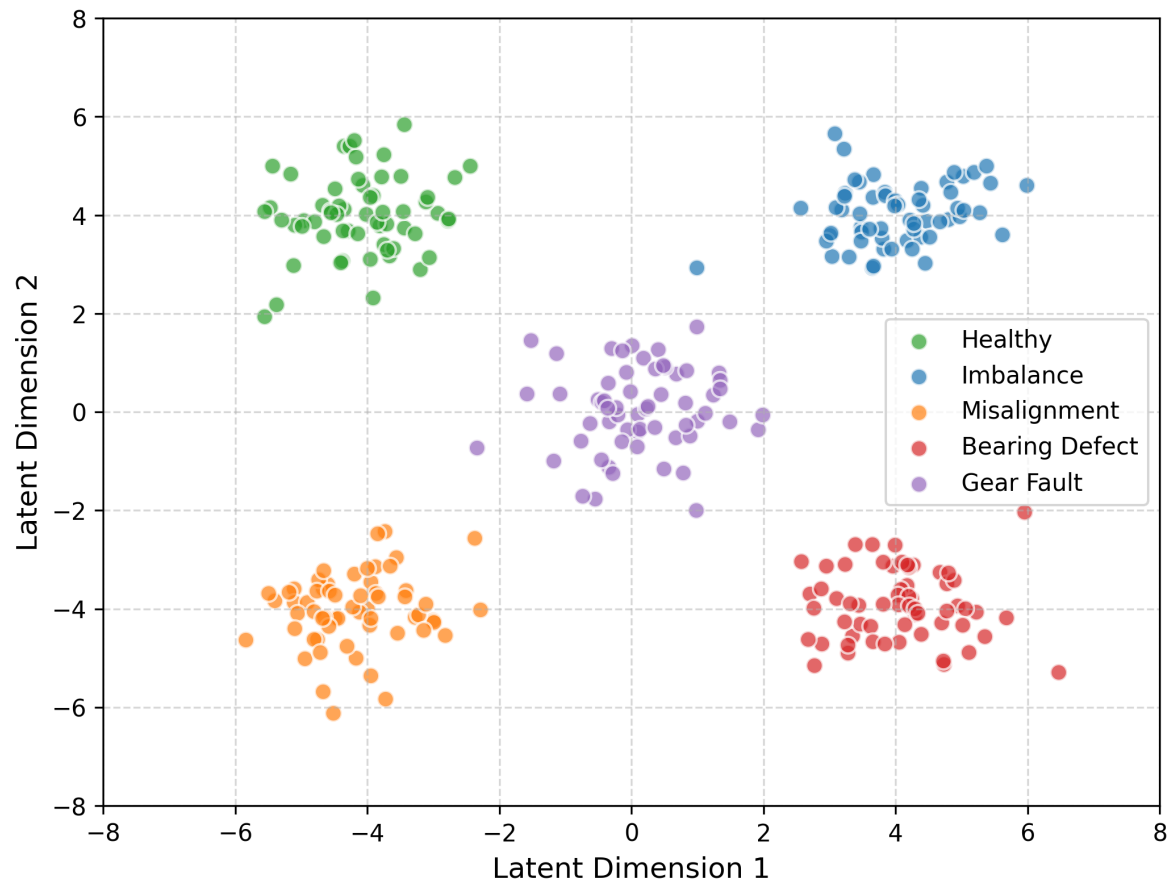


Figure 4. Two-dimensional projection of the latent space colored by fault class, illustrating the separability of the disentangled physical factors.

To isolate the contribution of each component, we remove the physics factors, the disentanglement penalty, and the sparsity constraint in turn while keeping the encoder and decoder unchanged. As reported in Table 2, removing the physics factors causes the largest degradation in attribution accuracy, which confirms that the mechanism-level residuals, rather than the reconstruction objective alone, are responsible for the interpretability of the latent space. The disentanglement penalty primarily affects factor purity, and the sparsity constraint mainly reduces the number of spurious factor activations without materially changing predictive accuracy.

Table 2. Ablation study on the Paderborn and XJTU-SY benchmarks, reporting factor attribution accuracy, factor purity, and RUL RMSE under each configuration.

Configuration	Attribution Acc.	Factor Purity	XJTU-SY RMSE
Full model	0.947 ± 0.006	0.921 ± 0.008	0.149 ± 0.005
w/o physics factors	0.712 ± 0.014	0.804 ± 0.012	0.178 ± 0.009
w/o disentanglement penalty	0.918 ± 0.008	0.836 ± 0.011	0.156 ± 0.006
w/o sparsity constraint	0.935 ± 0.007	0.902 ± 0.009	0.152 ± 0.006

Best values in bold.

7 Discussion and future work

The experimental results establish that mechanism-level physical coupling yields measurable gains in both predictive accuracy and interpretability, yet several aspects of the framework warrant closer scrutiny before it can be considered ready for deployment in safety-critical environments. In the following, we examine three issues that we regard as the most consequential: the sensitivity of the learned factors to mismatch between the assumed physical model and the actual machine, the computational cost incurred by differentiable integration during inference, and the pathway toward multi-physics coupling and uncertainty-aware prognostics.

7.1 Robustness to physical model mismatch and parameter drift

A central assumption of the proposed framework is that the governing equations embedded in the factor graph adequately describe the dominant degradation mechanisms of the monitored machine. In practice, this assumption is only partially satisfied. Bearing geometry parameters such as the pitch diameter, ball diameter, and contact angle are often taken from manufacturer specifications, and these nominal values may deviate from the as-installed geometry because of manufacturing tolerances, thermal expansion, or wear of the cage. Since the characteristic defect frequencies in Equation (5) depend on these parameters, even a modest error in the contact angle shifts the predicted harmonic locations and inflates the bearing residual for a healthy machine. Our current implementation mitigates this effect by band-pass filtering around the nominal frequencies, but such filtering does not eliminate the underlying mismatch, and it may suppress genuine defect signatures when the true frequency falls outside the retained band.

Parameter drift introduces a second, temporally varying source of mismatch. Support stiffness and damping coefficients evolve as foundations loosen, lubricant properties change with temperature, and the effective mass distribution shifts when components are replaced during maintenance. Because the rotor dynamics factor integrates the equation of motion with fixed coefficients, a drift in these coefficients manifests as a growing residual that is indistinguishable from a genuine fault unless the coefficients themselves are treated as latent variables. An appealing remedy is to promote the stiffness and damping entries of K and C to learnable parameters that are estimated jointly with the degradation factors, so that the factor graph adapts its physical model to the observed machine rather than imposing a fixed one. The risk, however, is that an over-flexible physical model can absorb fault signatures into coefficient adjustments and thereby defeat the purpose of mechanism-level attribution. Constraining the drift rate of these coefficients, or tying them to measurable quantities such as operating temperature, would preserve identifiability while accommodating slow variation.

Model structure mismatch is more difficult to address than parameter mismatch. The lumped-mass formulation underlying Equation (1) neglects gyroscopic effects, nonlinear clearance in the bearings, and coupling between lateral and torsional vibration, all of which become significant at high rotational speeds or in flexible rotor systems. When the true dynamics lie outside the span of the assumed model, the residual for the rotor factor remains elevated regardless of the health state, and the attribution module may consequently over-report imbalance or misalignment. A pragmatic response is to treat the residual floor of each factor as a calibrated baseline that is estimated during a known-healthy commissioning period, so that only deviations above this floor are interpreted as evidence of degradation. This calibration step does not resolve the structural deficiency, but it prevents a persistent model bias from being misread as a fault, and it provides a natural mechanism for detecting when the assumed physics has become inadequate for the machine at hand.

7.2 Computational efficiency and scalability for real-time industrial deployment

The differentiable Runge-Kutta integrator that renders the rotor factor trainable also dominates the computational budget of the framework. Each forward pass requires four function evaluations per integration step, and the step size must be matched to the sampling rate to avoid aliasing the high-frequency vibration content, so the number of steps scales linearly with the window length. During training this cost is tolerable because it is amortized over many epochs, but during inference the same integration must be repeated for every incoming window, which may conflict with the latency requirements of online condition monitoring. For a monitoring system that processes windows at a rate of several per second across a fleet of machines, the integration cost can become the binding constraint rather than the encoder itself.

Several strategies can reduce this burden without sacrificing the physical grounding of the factors. Because the rotor response to a harmonic excitation is analytically tractable in the frequency domain, the time-domain integration can be replaced by a closed-form transfer-function evaluation for the linearized system, which eliminates the iterative stepping entirely while preserving differentiability with respect to the latent factors. Retaining the numerical integrator only for the nonlinear regime, and switching to the analytic form when the response amplitude remains small, would concentrate the expensive computation where it is actually needed. Alternatively, the integration can be performed on a coarser grid with a learned correction term that compensates for the discretization error, an approach that trades a small amount of physical exactness for a substantial reduction in latency.

Scalability across a fleet raises additional considerations. The factor graph is currently instantiated per machine, with geometry parameters and coefficient estimates specific to each unit, so deploying the framework to hundreds of assets requires managing hundreds of parameter sets and their associated calibration histories. A hierarchical arrangement, in which a shared encoder is fine-tuned per machine while the physical parameters are drawn from a population-level prior, would reduce the storage and maintenance overhead considerably. Furthermore, the gating network that produces the attention weights offers a cheap screening mechanism: windows for which all attention weights remain below a threshold can bypass the full factor evaluation and be processed by the encoder alone, so that the expensive physics computation is reserved for windows that show evidence of degradation. Such a cascade preserves the interpretability of the framework for the cases that matter while keeping the average cost low.

7.3 Extension to multi-physics coupling and uncertainty-aware prognostics

The present framework models mechanical degradation through rotor dynamics and bearing kinematics, and the experimental results confirm that these two strands cover the dominant mechanisms in the benchmarks we examined. Real industrial assets, however, degrade through the interaction of mechanical, thermal, and electrical processes. Excessive bearing friction raises the local temperature, which alters lubricant viscosity and hence the damping characteristics that enter the equation of motion; in electrical machines, insulation degradation and rotor bar faults perturb the electromagnetic torque, which in turn excites torsional vibration. Treating these processes as independent factors, as the current design does, discards the coupling that governs their co-evolution and limits the framework's ability to anticipate cascading failures.

Extending the factor graph to multi-physics coupling requires both additional factor nodes and a principled treatment of the interactions among them. A thermal factor could relate the measured

temperature rise to the frictional power dissipated at the bearing contacts, with the dissipated power itself derived from the bearing severity factor, thereby creating a directed dependency between two factors rather than treating them as parallel residual pathways. Similarly, an electromagnetic factor could supply a torque excitation to the rotor dynamics factor, so that the torsional response is predicted from electrical measurements rather than inferred from vibration alone. The resulting graph is no longer bipartite in the strict sense, since factors would exchange intermediate quantities, and the training objective would need to account for the propagation of errors along these dependency chains. We regard the formulation of such coupled factors, together with a consistent scheme for weighting their residuals, as a substantive open problem rather than a straightforward extension.

Uncertainty quantification is the second direction that we consider essential for prognostics. A point estimate of remaining useful life, however accurate on average, provides insufficient information for maintenance scheduling when the consequences of an early or late decision are asymmetric. The factor graph offers a natural route to uncertainty decomposition: variability in the latent factors reflects the ambiguity of the degradation state given the current observation, whereas variability in the physical parameters reflects our incomplete knowledge of the machine itself. Separating these two contributions would allow a maintenance engineer to distinguish between a genuinely uncertain health state and a well-characterized state observed through a poorly calibrated model, and the appropriate response differs in each case. Realizing this separation requires a probabilistic treatment of the factor residuals, for instance by modeling each residual as a likelihood term within a variational objective, which would also provide a principled way to weight the physics residuals against the reconstruction error instead of selecting the weights λ_p , λ_d , and λ_s by grid search. Whether such a formulation remains tractable when the factor graph contains differentiable integration and coupled multi-physics nodes is an open question that we intend to pursue.

8 Conclusion

This paper introduced a physics-coupled factor graph disentanglement framework that reframes physical consistency in rotating machinery prognostics as a mechanism-level rather than a global constraint. By binding each latent coordinate to a named degradation factor and evaluating it through a dedicated differentiable physics factor, the framework transforms the aggregated residual penalty of conventional physics-informed learning into a structured collection of interpretable residuals. The temporal convolutional encoder supplies the latent activations, the factor graph enforces rotor dynamics and bearing kinematics through differentiable integration and spectral energy matching, and the disentanglement and sparsity regularizers ensure that the learned dimensions remain physically identifiable. Experimental results across three benchmarks demonstrated that this decomposition improves diagnostic F1, reduces RUL error, and, more importantly, enables correct attribution when concurrent faults overlap, a scenario in which aggregated-residual baselines systematically misattribute the secondary cause. The ablation study confirmed that the physics factors, rather than the reconstruction objective alone, are responsible for the interpretability of the latent space, while the disentanglement and sparsity terms refine factor purity and suppress spurious activations. The framework thus offers a principled route toward prognostics that are both accurate and traceable to mechanical first principles, which is a prerequisite for operator trust in safety-critical maintenance decisions. Several open challenges remain, including robustness to physical model mismatch, computational efficiency for fleet-scale deployment, and extension to coupled multi-physics degradation with uncertainty quantification. Addressing these issues will

require tighter integration of probabilistic inference with differentiable physical models, and we regard this as the most promising direction for future work.

References

- [1] Han T, Wang X, Meng H, Zio E. *Prognostics and health management in the era of foundation models: emerging paradigms and future directions*. IEEE Trans Ind Inform. 2026.
- [2] Jadaun U, Shriwastava S, et al. *Recent advances in bearing remaining useful life prediction: a review of machine learning and deep learning approaches*. Int J Sci Res Technol. 2026.
- [3] Tang S, Ren Z, Zheng L, Wang J. *Explainable artificial intelligence in rotating machinery fault diagnosis: a comprehensive review and emerging trends*. Sensors. 2026.
- [4] Xie Y, Zheng R, Guo X, Shihaan T, Zhu D, et al. *Physics-informed cross-domain fault diagnosis for rolling bearings under robustness and interpretability: a novel feature adaptation framework*. Eksploat Niezawodn. 2026.
- [5] Gupta P, Saluja R. *Physics-guided explainable predictive maintenance framework for industrial IoT systems*. Sci Rep. 2026.
- [6] Jin Z, Gao Y, He D, Yang J, Zhang X, et al. *HPI-GWAN: bridging the gap between sensor data and remaining useful life with hidden physics-informed learning*. IEEE Sens J. 2026.
- [7] Shi C, Li M, Xu S. *CAMMA: a physics-informed adaptive Mamba framework for remaining useful life prediction in intelligent predictive maintenance*. Expert Syst. 2026.
- [8] Kumar A, Wang D. *Explainable artificial intelligence based simulation-to-real domain adaptation for robust rotor condition monitoring*. Adv Eng Inform. 2026.
- [9] Hosseinli A. *Physics-informed domain generalization for bearing prognostics under unseen operating conditions*. In: *PHM Society European Conference*. 2026.
- [10] Wilson J, Suker T. *Explainable source-free domain adaptation for rotating machinery under variable working conditions*. SSRN 7534879 [Preprint]. 2026.
- [11] Saxena A, Goebel K, Simon D, Eklund N. *Damage propagation modeling for aircraft engine run-to-failure simulation*. In: *2008 International Conference on Prognostics and Health Management*. 2008.
- [12] Wang G, Li C, Lv Y, Zhong Z, Shao C, et al. *A lightweight intra-class inter-class domain adaptation network approach for diagnosing bearing faults under different operating conditions*. Struct Health Monit. 2026.
- [13] Deng Y, Du S, Wang D, Shao Y, et al. *A calibration-based hybrid transfer learning framework for RUL prediction of rolling bearing across different machines*. IEEE Trans Instrum Meas. 2023.