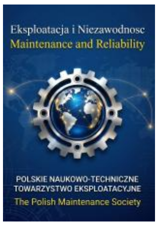

Eksploatacja i Niezawodnosc - Maintenance and Reliability

Volume 29 (2027), Issue 1

https://ein.org.pl

The Polish Maintenance Society

Physical-causal guided adaptive time-frequency and hypergraph co-evolutionary modeling method for industrial equipment monitoring

Ruoyan Ma[1,*]

[1] College of Computing, City University of Hong Kong, Hong Kong Special Administrative Region, China
* Corresponding Author: alice_kris@foxmail.com

Received: 4 April 2026
Revised: 22 June 2026
Accepted: 28 July 2026
Online: 13 August 2026

Abstract

Accurate monitoring of industrial equipment is important for production safety and reliability. This paper proposes a physical-causal guided adaptive time-frequency and hypergraph co-evolutionary modeling method. The method constructs neural network representations with parameterized basis functions, enabling adaptive time-frequency decomposition with energy conservation and modal orthogonality constraints. Differentiable physical projection operators ensure hard constraint satisfaction during training. A transfer entropy-based causal discovery mechanism converts causal relationships into hyperedge generation rules for dynamic hypergraph construction. Bidirectional cross-attention mechanisms achieve collaborative updates between time-frequency and hypergraph features. Experiments achieve 98.73% fault diagnosis accuracy on CWRU dataset, 91.77% cross-working-condition transfer accuracy, 82.56% accuracy at -6dB SNR, and 82.4% cross-dataset generalization accuracy on unseen equipment types.

Keywords: *physical-causal constraints, adaptive time-frequency decomposition, hypergraph neural network, co-evolutionary mechanism, industrial equipment monitoring*

Article citation:
Ma R, Physical-causal guided adaptive time-frequency and hypergraph co-evolutionary modeling method for industrial equipment monitoring, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2027: 29(1) http://doi.org/10.17531/ein/226533

Highlights

- A physical causal framework jointly learns adaptive time frequency representations and dynamic causal hypergraphs for equipment monitoring across operating conditions.
- Differentiable physical projections enforce energy conservation, frequency nonnegativity, and modal orthogonality during adaptive basis function learning for physically plausible decomposition.
- Transfer entropy hypergraphs and bidirectional cross attention improve diagnosis, noise robustness, condition transfer, and cross dataset generalization across different equipment.

1. Introduction

With Industrial IoT development, intelligent monitoring of industrial equipment has become essential for safe production. Modern industrial sites deploy numerous sensors collecting multimodal signals. However, monitoring data exhibit high dimensionality, strong coupling, and non-stationarity characteristics.

Current research follows signal processing and deep learning routes. Hou et al. [1] proposed differential modal decomposition. Niu et al. [2] constructed multi-task bearing fault diagnosis based on deep residual CNNs. Wang et al. [3] introduced spatiotemporal graph neural networks for multi-sensor fusion. However, these methods share several common limitations. Fixed basis functions are unable to adapt to time-varying signal characteristics, binary graph edges are insufficient to characterize multi-body high-order interactions, and purely data-driven approaches tend to generalize poorly when training samples are limited.

Wang et al. [4] proposed multilayer wavelet attention CNNs but lacked physical constraint guidance. Wang et al. [5] proposed tensor hypergraph neural networks, and Gao et al. [6] developed general hypergraph frameworks. However, most hypergraph methods use static prior knowledge without considering dynamic evolution. Miele et al. [7] proposed physics-informed probabilistic deep networks. Liu et al. [8] proposed causal attention graph neural networks.

This paper proposes a physical-causal guided adaptive time-frequency and hypergraph co-evolutionary method, establishing bidirectional coupling between time-frequency extraction and hypergraph modeling with physical-causal constraints as the connecting link.

Main contributions include:

(1) A co-evolutionary framework with learnable parameterized basis functions, causally-driven dynamic hypergraphs, and bidirectional cross-attention.

(2) Physical constraint regularized adaptive time-frequency basis function learning.

(3) Causally-driven dynamic hypergraph modeling with transfer entropy-based causal discovery.

2. Related work

2.1. Adaptive time-frequency analysis methods

Time-frequency analysis maps signals to time-frequency planes revealing local frequency characteristics. Traditional methods face difficulties optimizing time and frequency resolution simultaneously. Miao et al. [9] proposed feature mode decomposition. Chen et al. [10] designed anti-noise adaptive feature mode decomposition. Li et al. [11] designed adaptive synchrosqueezing transform. Chen et al. [12] proposed impact time-domain adaptive decomposition. These methods have made contributions to adaptive signal decomposition, but they generally rely on predefined mathematical basis functions and do not incorporate physical constraint guidance into the decomposition process.

2.2. Hypergraph neural networks

Hypergraphs allow one hyperedge to connect multiple nodes simultaneously. Yan et al. [13] proposed multi-resolution hypergraph neural networks. Zhao et al. [14] designed model-assisted multi-source fusion hypergraph CNNs. Chen et al. [15] proposed interaction-aware graph neural networks. Wang et al. [16] proposed hypergraph-based coupled fault diagnosis. Qin et al. [17] designed fuzzy cognitive hypergraph models. A common limitation among these works is the reliance on static prior knowledge to define hyperedge structures, without mechanisms to update the topology in response to changing operating conditions or fault evolution.

2.3. Physical constraints and causal analysis

Integrating physical priors into deep learning improves interpretability. Ni et al. [18] proposed physics-informed residual networks. Xu et al. [19] designed physics-informed probabilistic networks. Liu et al. [20] proposed causal intervention graph neural networks. Zhang et al. [21] designed Granger causality-based graph neural networks. Wang et al. [22] proposed causal attention graph neural networks. A limitation shared by many physics-informed approaches is that physical knowledge enters the optimization as soft penalty terms in the loss function, which cannot guarantee that the learned representations strictly satisfy the underlying physical constraints throughout training.

3. Overall framework

Time-frequency feature extraction and multi-measurement-point relationship modeling are two core elements for equipment monitoring. This paper constructs a physical-causal guided framework establishing bidirectional coupling between them. The overall framework is illustrated in Figure 1. The framework contains three collaborating modules. The first module extracts time-frequency features through neural networks with parameterized basis functions, with physical plausibility ensured through energy conservation, frequency non-negativity, and modal orthogonality constraints; differentiable physical projection operators based on Gram-Schmidt orthogonalization transform these constraints from soft penalties into hard embeddings during training. The second module employs transfer entropy to quantify causal relationship strength among measurement points and converts the results into hyperedge connection rules for dynamic hypergraph construction; the causally-driven hyperedge generation captures high-order interaction patterns that binary graph edges are not designed to represent. The third module applies bidirectional cross-attention mechanisms to compute correlation matrices between time-frequency and hypergraph features, enabling mutual enhancement through iterative updates. The fused representations encode both time-frequency characteristics and causal structural information for final equipment condition monitoring output.

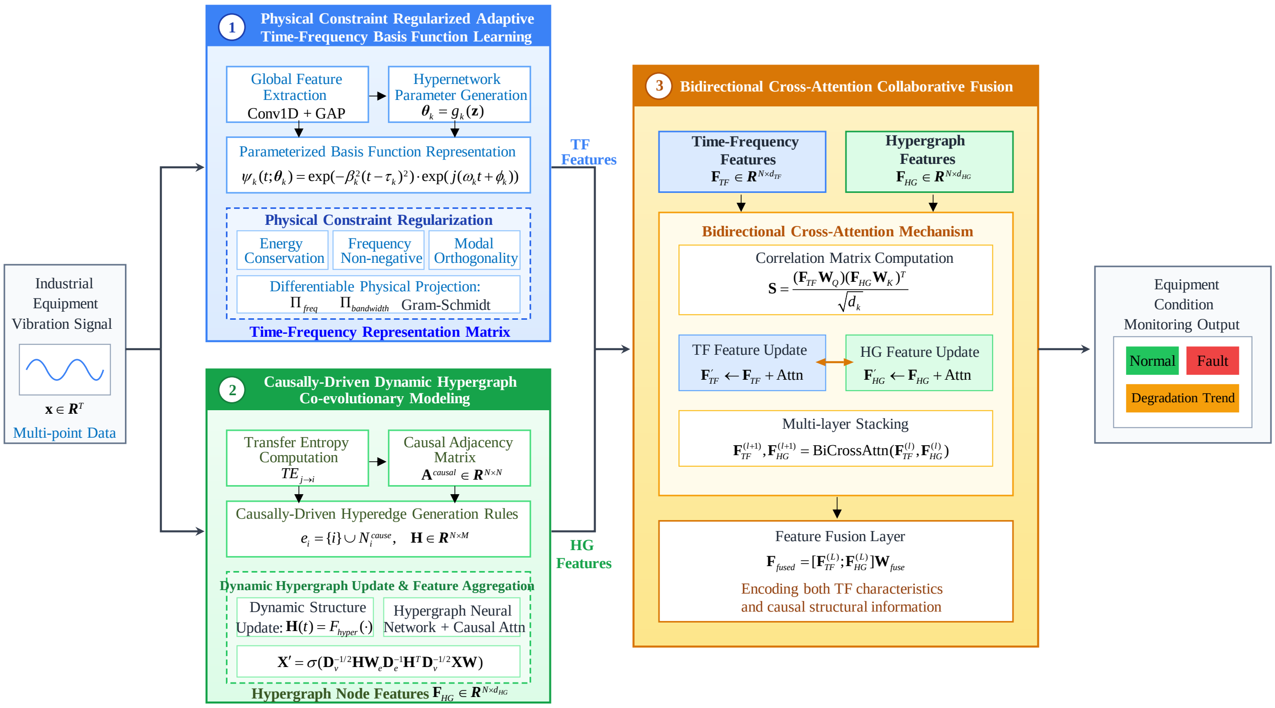

Figure 1. Physical-causal guided adaptive time-frequency and hypergraph co-evolutionary modeling method for industrial equipment monitoring.

4. Physical constraint regularized adaptive time-frequency basis function learning

Industrial equipment vibration signals have non-stationary characteristics with fault features manifesting as superposition of transient impacts and modulation patterns. Traditional time-frequency methods use predefined basis functions whose characteristics are determined by mathematical definitions rather than signals themselves.

4.1. Neural network representation with parameterized basis functions

Given input signal $\boldsymbol{x} \in R^T$, global features are obtained through $\boldsymbol{z} = f_{global}(\boldsymbol{x}) = \text{GAP}(\text{Conv1D}(\mathbf{x}))$. Basis function parameters are generated through hypernetworks as $\theta_k = g_k(\boldsymbol{z})$, containing center frequency ω_k, bandwidth β_k, and phase ϕ_k. The adaptive basis function is $\psi_k(t;\boldsymbol{\theta}_k) = exp(-\beta_k^2(t-\tau_k)^2)\, exp(j(\omega_k t + \phi_k))$.

Time-frequency coefficients are $c_k(\tau) = \langle \boldsymbol{x}, \psi_k(\cdot - \tau; \boldsymbol{\theta}_k)\rangle = \int_{-\infty}^{+\infty} x(t)\psi_k^*(t-\tau;\boldsymbol{\theta}_k)dt$. All components form the time-frequency matrix $\boldsymbol{C} \in C^{K\times T}$.

4.2. Physical constraint regularization

Energy conservation constraint requires $L_{energy} = \left|\|\boldsymbol{x}\|_2^2 - \sum_{k=1}^{K}\|c_k\|_2^2\right|$, ensuring decomposition does not create or lose signal energy.

Frequency non-negativity constraint is $L_{freq} = \sum_{k=1}^{K} max(0, -\omega_k) + \sum_{k=1}^{K} max(0, \omega_k - \omega_{max})$, preventing center frequencies from exceeding physically interpretable ranges.

Modal orthogonality constraint is $L_{orth} = \sum_{i=1}^{K}\sum_{j\neq i}^{K} \left|\langle\psi_i, \psi_j\rangle\right|^2$, encouraging different basis functions to maintain approximate orthogonality.

4.3. Differentiable physical projection operators

Projection operators project intermediate results to feasible domains during forward propagation. For frequency non-negativity, $\Pi_{freq}(\omega) = \text{softplus}(\omega) = log(1 + exp(\omega))$. For bandwidth positive definiteness, $\Pi_{bandwidth}(\beta) = \varepsilon + \text{softplus}(\beta)$.

For modal orthogonality, soft orthogonalization uses $\hat{\psi}_k = (1-\alpha)\psi_k + \alpha\psi_k^{\perp}$, where $\psi_k^{\perp} = \psi_k - \sum_{j=1}^{k-1}\frac{\langle\psi_k, \psi_j^{\perp}\rangle}{\langle\psi_j^{\perp}, \psi_j^{\perp}\rangle}\psi_j^{\perp}$.

The total loss function is $L_{TF} = L_{recon} + \lambda_1 L_{energy} + \lambda_2 L_{freq} + \lambda_3 L_{orth}$, where $L_{recon} = ||\boldsymbol{x} - \sum_{k=1}^{K}\text{Re}(c_k * \psi_k)||_2^2$.

The Gram-Schmidt orthogonalization procedure described above is applied on a per-sample basis during each forward pass. For a given input sample, the hypernetwork generates initial basis function parameters, and the iterative projection is then carried out over the K basis functions associated with that sample alone. Because the procedure operates independently for each sample without accumulating any state across the mini-batch, it is fully compatible with standard stochastic gradient training. Samples within the same mini-batch undergo orthogonalization independently of one another, and the process admits efficient parallelization through vectorization across the batch dimension.

5. Causally-driven dynamic hypergraph co-evolutionary modeling method

5.1. Transfer entropy-based causal discovery mechanism

Transfer entropy from measurement point j to i is $TE_{j\to i} = \sum p\,(x_i^{(t+1)}, x_i^{(t)}, x_j^{(t)})\, log \frac{p(x_i^{(t+1)}|x_i^{(t)}, x_j^{(t)})}{p(x_i^{(t+1)}|x_i^{(t)})}$, measuring the additional contribution of measurement point j's history to predicting measurement point i's future state. The causal adjacency matrix $\boldsymbol{A}^{causal} \in R^{N\times N}$ has elements $A_{ij}^{causal} = TE_{j\to i}$, reflecting causal relationship directionality.

5.2. Causally-driven hyperedge generation rules

For each measurement point i, causal neighborhoods are $N_i^{cause} = \{j: A_{ji}^{causal} > \theta_{cause}\}$. With measurement point i as center, all nodes in its causal neighborhood form hyperedge $e_i = \{i\} \cup N_i^{cause}$. The incidence matrix $\boldsymbol{H} \in R^{N\times M}$ is defined as $H_{ve} = 1$ if node v belongs to hyperedge e.

Dynamic hypergraph update uses $\boldsymbol{H}(t) = F_{hyper}(\boldsymbol{A}^{causal}(t), \theta_{cause}(t))$, where $\theta_{cause}(t) = \mu_A(t) + \gamma\sigma_A(t)$ is automatically determined from causal strength distribution statistics.

5.3. Hypergraph neural network feature aggregation

Hyperedge features are computed as $\boldsymbol{Y}_e = \frac{1}{|e|}\sum_{v\in e} \boldsymbol{X}_v \boldsymbol{W}_{n2e}$. Node feature updates are $\boldsymbol{X}_{v'} = \sigma\left(\sum_{e:v\in e} \frac{1}{|e|} \boldsymbol{Y}_e \boldsymbol{W}_{e2n}\right)$. Unified representation is $\boldsymbol{X}' = \sigma(\boldsymbol{D}_v^{-1/2} \boldsymbol{H} \boldsymbol{W}_e \boldsymbol{D}_e^{-1} \boldsymbol{H}^T \boldsymbol{D}_v^{-1/2} \boldsymbol{X}\boldsymbol{W})$. Causal attention weights node contributions within hyperedges as $\alpha_{v,e} = \frac{exp(A_{v,center(e)}^{causal})}{\sum_{u\in e} exp\, \Sigma_{u,center(e)}^{causal}}$.

5.4. Bidirectional cross-attention collaborative mechanism

Given time-frequency feature matrix $\boldsymbol{F}_{TF} \in R^{N\times d_{TF}}$ and hypergraph feature matrix $\boldsymbol{F}_{HG} \in R^{N\times d_{HG}}$, correlation matrix is $\boldsymbol{S} = \frac{\boldsymbol{F}_{TF}\boldsymbol{W}_Q(\boldsymbol{F}_{HG}\boldsymbol{W}_K)^T}{\sqrt{d_k}}$.

Time-frequency feature updates are $\boldsymbol{F}_{TF'} = \boldsymbol{F}_{TF} + \mathrm{softmax}(\mathbf{S})\mathbf{F}_{HG}\boldsymbol{W}_V^{HG}$. Hypergraph feature updates are $\boldsymbol{F}_{HG'} = \boldsymbol{F}_{HG} + \mathrm{softmax}(\mathbf{S}^T)\boldsymbol{F}_{TF}\boldsymbol{W}_V^{TF}$.

After multi-layer collaborative updates, fused features are $\boldsymbol{F}_{fused} = [\boldsymbol{F}_{TF}^{(L)}; \boldsymbol{F}_{HG}^{(L)}]\boldsymbol{W}_{fuse}$, encoding both time-frequency characteristics and causal structural information.

6. Experimental verification and analysis

6.1. Dataset and experimental setup

This paper uses the CWRU bearing dataset [23] for verification. The experimental platform consists of a 2-horsepower motor, torque sensor, power meter, and electronic controller. The bearing model is SKF 6205-2RS with 12 kHz sampling frequency. Fault types include ball fault, inner race fault, and outer race fault with diameters of 0.007, 0.014, and 0.021 inches. Combined with normal state, 10 fault patterns are formed. Motor speeds are 1797, 1772, 1750, and 1730 rpm corresponding to loads of 0, 1, 2, and 3 hp. Each sample extracts 2048 sampling points. Training and test sets are divided in 7:3 ratio with 5 repetitions.

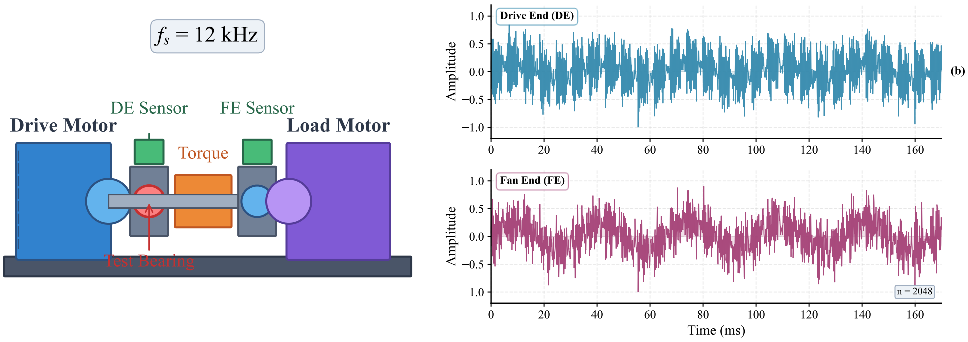

Figure 2. CWRU bearing dataset experimental platform and signal acquisition schematic. The left side shows the test rig structure including drive motor, test bearing, torque sensor, and load motor. The right side displays vibration signal waveforms from drive end and fan end measurement points with 12 kHz sampling frequency and 2048 sampling points per sample.

Vibration signals from both drive end and fan end accelerometers are used. Network training uses Adam optimizer with learning rate 0.001, batch size 64, and maximum 200 epochs. The number of adaptive basis functions is 8, causal threshold γ is 1.0, and cross-attention layers are 3. Experiments are completed on NVIDIA RTX 3090. The experimental platform and typical signal waveforms are shown in Figure 2.

6.2. Overall fault diagnosis performance comparison

Multiple methods are selected for comparison: VMD [24], WDCNN [25], CNN [26], DRSN [27], MRF-GCN [28], TFT [29], and Attention-CNN [30]. For VMD [24], vibration modes are first decomposed using variational mode decomposition, and a fixed set of time-domain and frequency-domain statistical features—including root mean square, kurtosis, skewness, peak-to-peak value, and spectral centroid—are extracted from each mode. These features are concatenated into a feature vector and fed into a support vector machine classifier with a radial basis function kernel. The mode number K and the penalty coefficient C are jointly determined through five-fold cross-validation on the training set, and the bandwidth constraint parameter is fixed at 2000 as recommended in the original VMD formulation [24]. The reported accuracy of 92.35% corresponds to this complete VMD+SVM pipeline.

The proposed method achieves 98.73±0.21% accuracy, outperforming VMD (92.35%), CNN (95.67%), WDCNN (96.82%), DRSN (97.24%), MRF-GCN (97.56%), Attention-CNN (97.83%), and TFT (98.12%). The proposed method also achieves the highest AUC of 0.9987, compared with TFT (0.9956), DRSN (0.9912), MRF-GCN (0.9878), WDCNN (0.9845), and CNN (0.9723). Compared with VMD, accuracy improves by 6.38 percentage points. Compared with MRF-GCN, accuracy improves by 1.17 percentage points, indicating that hypergraph representations offer advantages in high-order relationship characterization.

The confusion matrix for 10-class fault recognition is presented in Figure 3. Diagonal elements show high values, indicating accurate identification of all fault types. Misclassified samples mainly concentrate between different damage degrees of the same fault type, conforming to the observation that adjacent damage degrees tend to produce more similar feature representations.

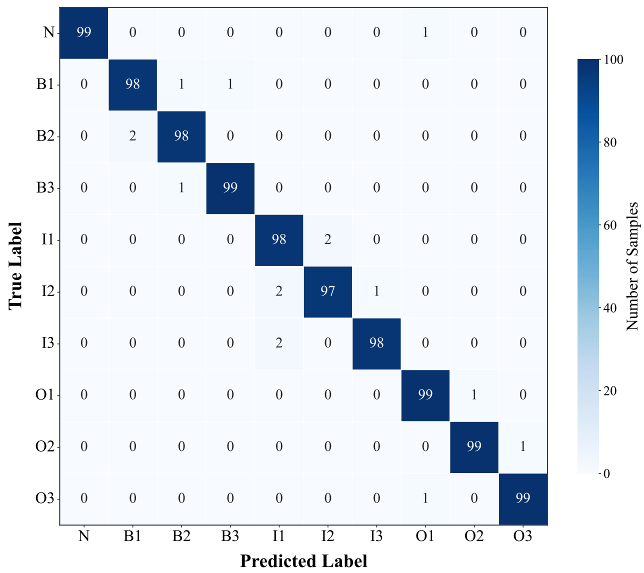

Figure 3. Confusion matrix of the proposed method on CWRU dataset. The horizontal axis represents predicted labels and vertical axis represents true labels. Diagonal elements indicate correctly classified samples with color depth reflecting magnitude. Categories 0-9 correspond to normal, ball fault 0.007"/0.014"/0.021", inner race fault 0.007"/0.014"/0.021", and outer race fault 0.007"/0.014"/0.021".

6.3. Ablation experiments

Systematic ablation experiments verify effectiveness of core components. The full model achieves 98.73% accuracy. Removing physical constraints causes 0.88 percentage points decrease. Removing causal discovery causes 1.31 percentage points decrease, the largest among all configurations, indicating that transfer entropy-based causal analysis plays an important role in capturing information flow directions among measurement points. Using fixed basis functions causes 1.20 percentage points decrease, validating advantages of adaptive basis function learning. Replacing the hypergraph with an ordinary graph causes a 1.05 percentage point decrease, indicating that high-order relationship modeling contributes measurably to diagnostic performance. Removing cross-attention causes 0.82 percentage points decrease.

As shown in Figure 4, the accuracy distribution across fault types reveals that the full model shows most complete contour with high values in all directions. After removing causal discovery, performance decreases notably on inner race and ball fault recognition, possibly because these fault types exhibit complex causal transfer patterns across multiple measurement points.

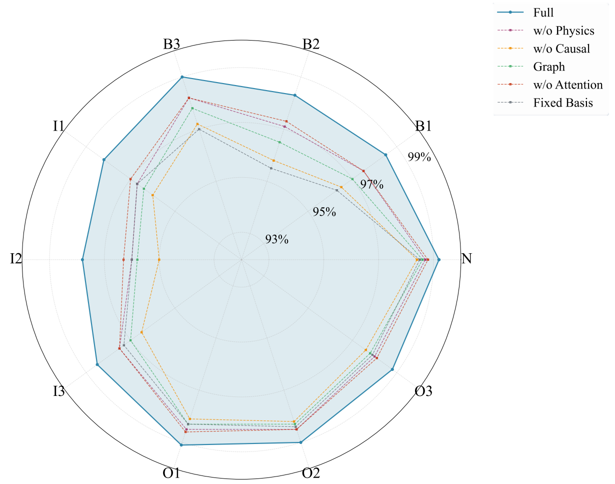

Figure 4. Radar chart of recognition accuracy for different ablation configurations across fault types. Each axis represents a fault type with values increasing outward. Different colored curves correspond to six configurations. Larger radar chart area indicates better overall performance.

6.4. Variable working condition transfer experiments

Cross-working-condition transfer experiments are designed with working conditions A, B, C, D corresponding to 0, 1, 2, 3 hp loads. The proposed method achieves 91.77% average accuracy, improving by 2.93 percentage points over TFT (88.84%). For A→D transfer with maximum working condition span from 0 to 3 hp, the proposed method maintains 88.72% while VMD achieves only 71.23%. Physical constraints ensure learned time-frequency features conform to energy conservation principles that remain unchanged under different working conditions. Causally-driven hypergraph structures focus on causal transfer relationships rather than signal amplitudes, and causal relationships maintain working condition stability at the fault mechanism level.

The performance comparison through heatmap is illustrated in Figure 5. The proposed method shows overall deeper colors and uniform distribution, indicating high and stable performance across transfer directions. The TFT method shows deeper colors near the diagonal but lighter colors in distant regions, indicating sensitivity to working condition changes.

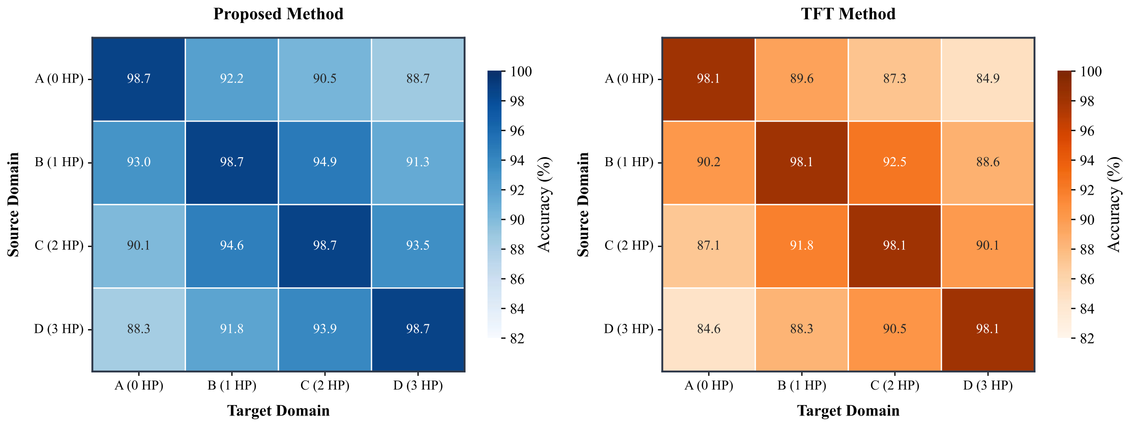

Figure 5. Accuracy heatmap of different methods under various working condition transfer scenarios. The horizontal axis is target domain working condition and vertical axis is source domain. Color depth represents accuracy level. Left side shows the proposed method and right side shows TFT method.

6.5. Cross-dataset generalization experiments

To verify the generalization capability of the proposed method across different equipment types and operating conditions, cross-dataset transfer experiments are conducted on four additional public datasets: PHM 2012 bearing dataset [31], MFPT dataset [32], Paderborn University (PU) dataset [33], and SEU gearbox dataset [34]. The PHM 2012 dataset contains run-to-failure bearing degradation data with 2560 Hz sampling frequency [31]. The MFPT dataset includes bearing fault data collected under real industrial noise conditions with 48828 Hz sampling frequency [32]. The PU dataset provides vibration signals from bearings with artificial and real damages under various operating conditions [33]. The SEU gearbox dataset contains gearbox fault signals, which is used to validate method transferability to different equipment types beyond bearings [34].

The model trained on CWRU dataset is directly applied to these datasets without fine-tuning to evaluate zero-shot transfer capability. For fair comparison, all signals are resampled to 12 kHz and segmented into 2048-point samples consistent with CWRU preprocessing. The cross-dataset generalization performance is illustrated in Figure 6. The proposed method achieves 89.3%, 87.2%, 85.7%, and 82.4% accuracy on PHM, MFPT, PU, and SEU datasets respectively, outperforming TFT [29] by 4.5, 5.2, 6.3, and 7.8 percentage points. Compared with MRF-GCN [28], the proposed method shows 5.8, 6.4, 7.2, and 8.5 percentage points improvement. The performance gap between the proposed method and comparison methods increases on more challenging datasets, particularly on SEU gearbox dataset where equipment type differs from the source domain.

The transfer from CWRU bearing data to SEU gearbox data involves two levels of domain shift. At the device level, bearing fault signatures are characterized by impulsive excitations at characteristic frequencies (BPFI, BPFO, BSF) determined by rolling element geometry, whereas gearbox fault signatures are governed by gear mesh frequency harmonics and modulation sidebands arising from shaft rotation. These two excitation mechanisms produce time-frequency patterns that differ in their structural organization. At the signal statistics level, gearbox vibration signals typically contain more complex multi-source coupling components, leading to differences in spectral envelope shape and degree of non-stationarity compared with bearing signals.

The proposed method retains a relatively competitive level of performance in this scenario through two complementary aspects of its design. The physical constraints imposed on time-frequency decomposition— energy conservation, frequency non-negativity, and modal orthogonality—are properties of physical signals that hold regardless of the source equipment type. As a result, the learned basis functions conform to physically interpretable frequency structures independent of the source domain. The transfer entropy-based hypergraph construction focuses on inter-sensor causal flow patterns rather than signal amplitude statistics. Under fault conditions, the tendency for drive-end sensors to exert stronger causal influence on downstream sensors reflects the mechanical transmission path. While the quantitative magnitude of these causal strengths differs between bearing and gearbox systems, the directional structure of the causal propagation pattern retains a consistent character that supports cross-equipment transferability.

Physical constraints ensure that learned time-frequency features conform to universal energy conservation principles that remain valid across different equipment. Causally-driven hypergraph structures focus on fundamental causal transfer patterns rather than dataset-specific amplitude characteristics, enabling better cross dataset generalization.

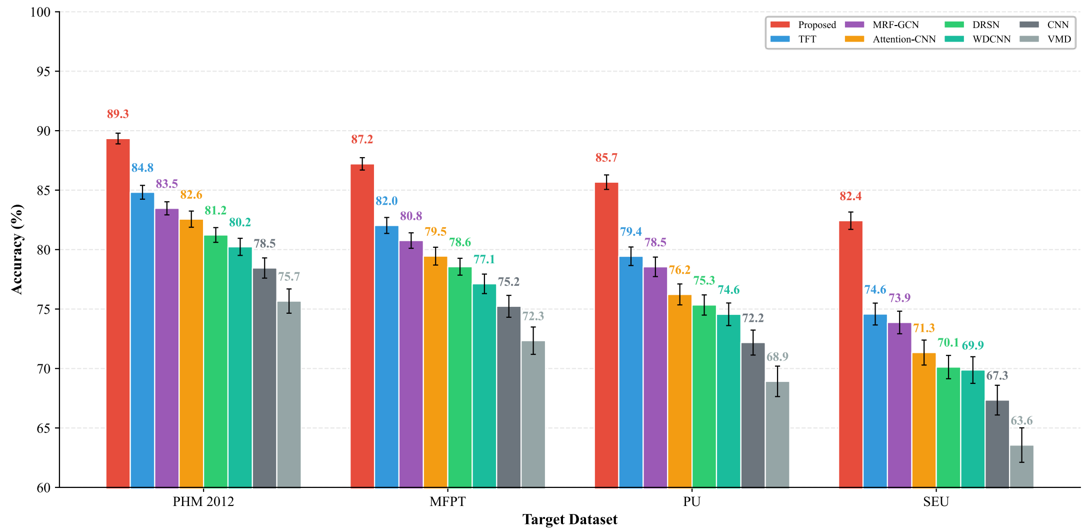

Figure 6. Cross-dataset generalization performance comparison. The horizontal axis represents four target datasets (PHM 2012 [31], MFPT [32], PU [33], and SEU [34]) and the vertical axis represents diagnostic accuracy. Different colored bars correspond to eight methods. The model is trained on CWRU dataset and directly tested on target datasets without fine-tuning. Error bars represent standard deviations from five repeated experiments.

6.6. Noise robustness analysis

Gaussian white noise of different intensities is added to test data with SNR of -6dB to 6dB. The diagnostic accuracy curves under different noise intensities are shown in Figure 7. At -6dB, the proposed method achieves 82.56% while CNN [26] only achieves 62.45%, a difference of 20.11 percentage points. At 6 dB, the proposed method reaches 98.73%. Physical constraints make the model learn features that conform to physical laws rather than noise components. Causally-driven hypergraph structures focus on causal relationships that have greater stability under noise perturbations.

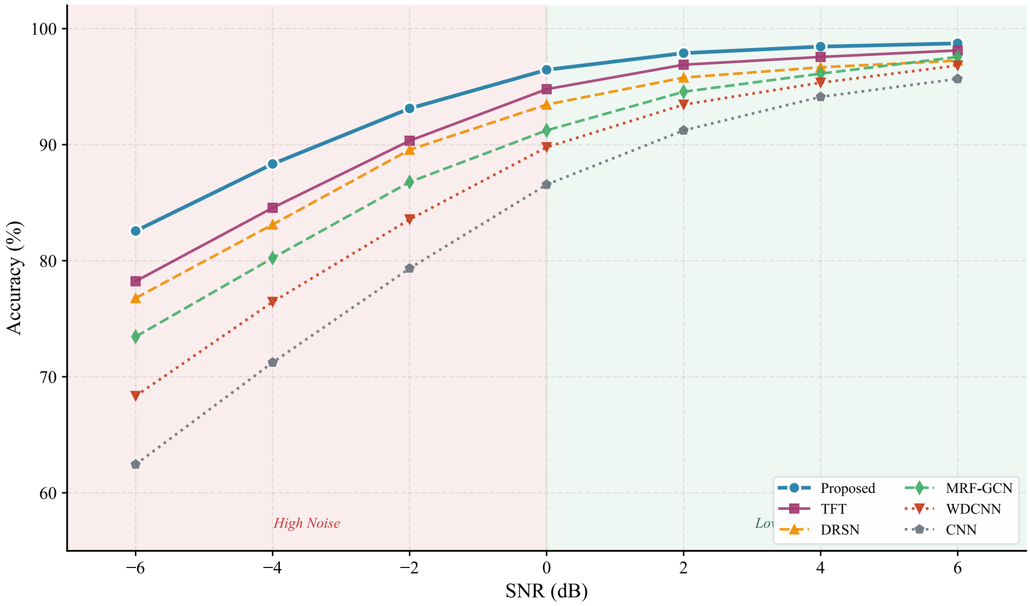

Figure 7. Diagnostic accuracy curves of different methods under various SNR conditions. The horizontal axis shows SNR levels ranging from -6dB to 6dB and vertical axis shows diagnostic accuracy. Different colored curves correspond to six methods including Proposed, TFT [29], DRSN [27], MRF-GCN [28], WDCNN [25], and CNN [26].

6.7. Time-frequency feature visualization analysis

To understand performance in time-frequency feature extraction, visualization analysis is conducted. As shown in Figure 8, the proposed adaptive basis functions achieve better time-frequency resolution balance near bearing characteristic frequencies compared with traditional Morlet wavelet. Traditional Morlet wavelet shows a fixed change pattern along the frequency axis, while the proposed method shows adaptive matching characteristics to fault characteristic frequency bands.

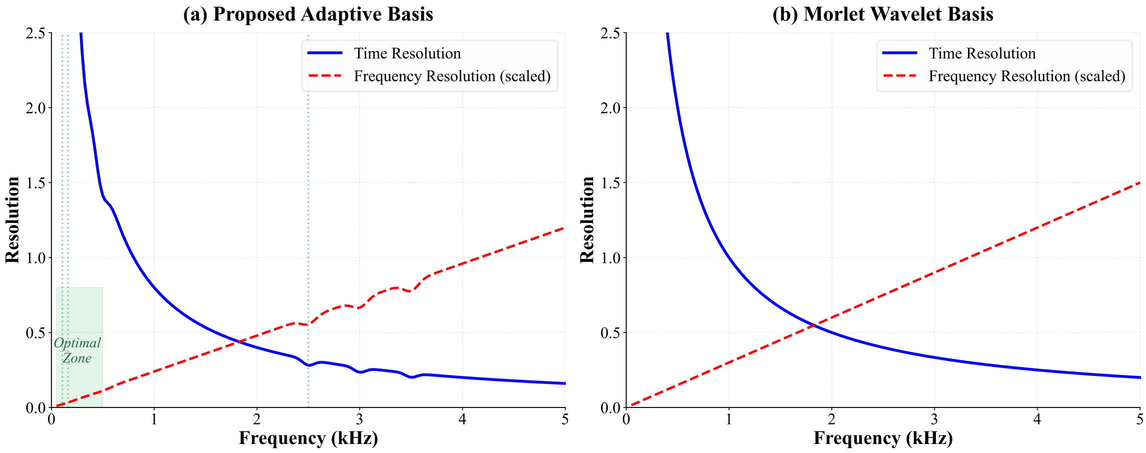

Figure 8. Time-frequency resolution comparison between adaptive and fixed basis functions. The left side shows the proposed adaptive basis function resolution distribution and right side shows traditional Morlet wavelet. The horizontal axis is frequency and vertical axis is time resolution.

6.8. Hypergraph structure and causal relationship visualization

The inter-measurement-point causal adjacency matrix is visualized in Figure 9. The asymmetric structure reflects information flow directionality. Drive end signals have higher causal influence on fan end signals than the reverse direction, which is consistent with the physical structure where the fault source is at the drive end bearing and vibration signals propagate to the fan end through the shaft system.

Figure 10 compares the learned hypergraph structures between normal and fault states. The normal state shows a relatively sparse hypergraph with low causal association strength. The fault state shows a more complex topology with drive end measurement points showing higher node importance as fault sources. High-order hyperedges connecting multiple measurement points increase in number and weight, reflecting more active vibration energy transfer among measurement points under fault conditions.

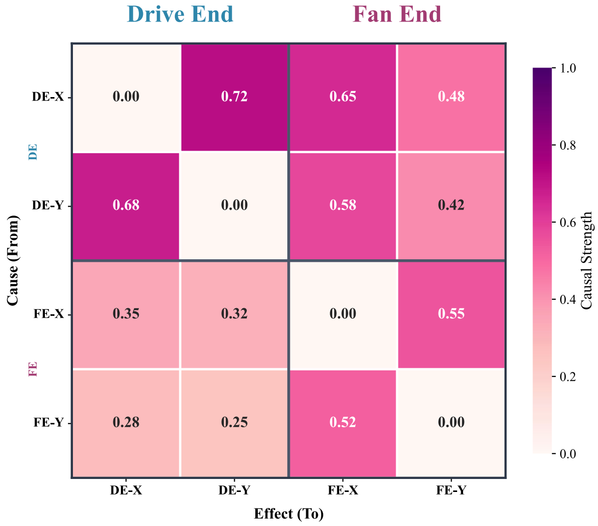

Figure 9. Heatmap visualization of causal adjacency matrix. Horizontal and vertical axes correspond to drive end X, drive end Y, fan end X, and fan end Y measurement points. Matrix elements represent causal strength from transfer entropy calculation. Deeper colors indicate stronger causal influence. Asymmetric matrix structure reflects causal relationship directionality.

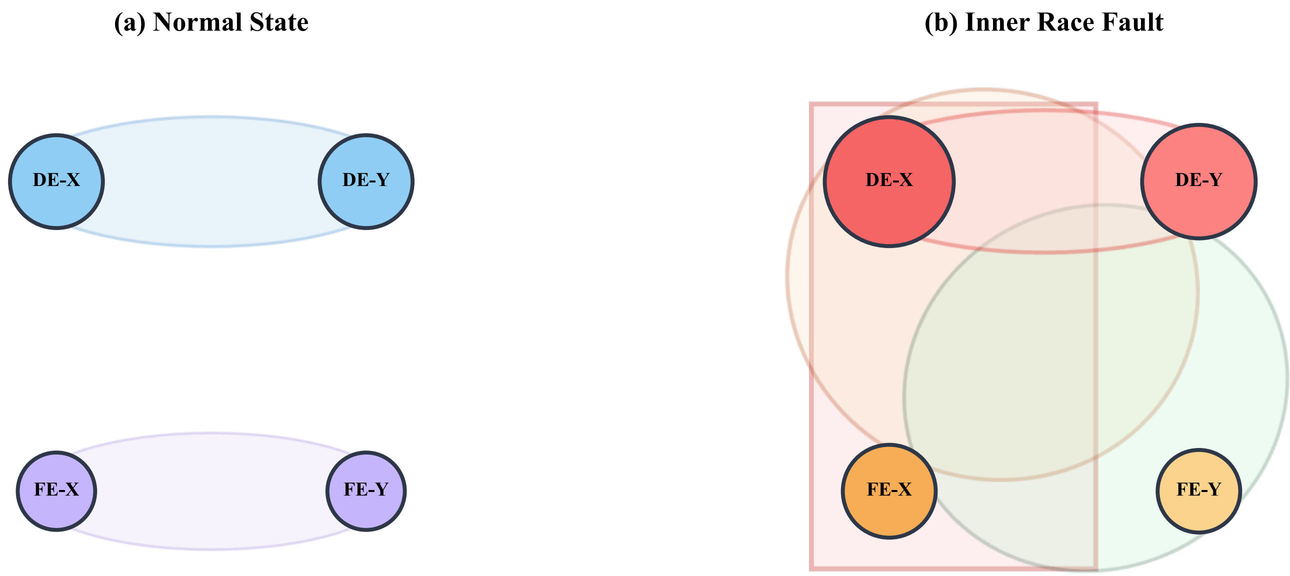

Figure 10. Comparative visualization of hypergraph structures under normal and fault states. The left side shows normal state hypergraph structure and right side shows inner race fault state. Nodes represent measurement points. Hyperedges are represented by closed curves surrounding multiple nodes. Curve thickness reflects hyperedge weight. Node size reflects importance in hypergraph.

6.9. Physical constraint effectiveness verification

Physical constraints are important components of the proposed method. As illustrated in Figure 11, with physical constraints, energy reconstruction error remains at approximately 0.8%. Without constraints, the error fluctuates with some samples exceeding 15%. Physical constraints effectively control energy reconstruction error within a reasonable range through energy conservation regularization terms in the loss function.

Figure 12 compares the orthogonality of the learned basis functions. With orthogonality constraint, inner product absolute values between basis functions are generally small, with average non-diagonal element approximately 0.05. Without constraint, inner product values are larger with some exceeding 0.3, indicating modal aliasing among components.

The correspondence between learned basis function center frequencies and bearing theoretical characteristic frequencies is shown in Figure 13. Distribution peaks highly match rotation frequency f_r, BSF, BPFO, BPFI, and resonance frequencies, indicating that the model learns frequency components related to fault mechanisms through a data-driven approach without requiring pre-specified characteristic frequencies.

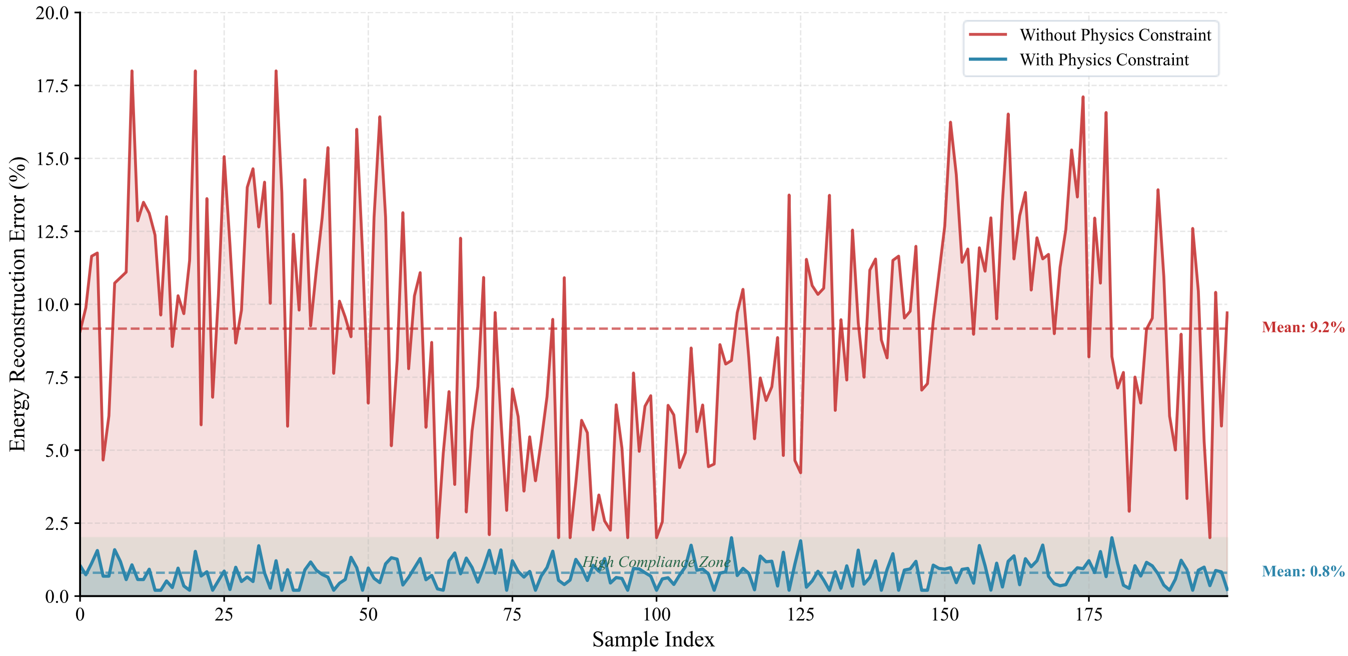

Figure 11. Analysis of energy conservation constraint satisfaction degree. The horizontal axis is sample index and vertical axis is energy reconstruction error ratio. Two curves correspond to configurations with and without physical constraints. Error closer to zero indicates higher satisfaction degree of energy conservation constraint.

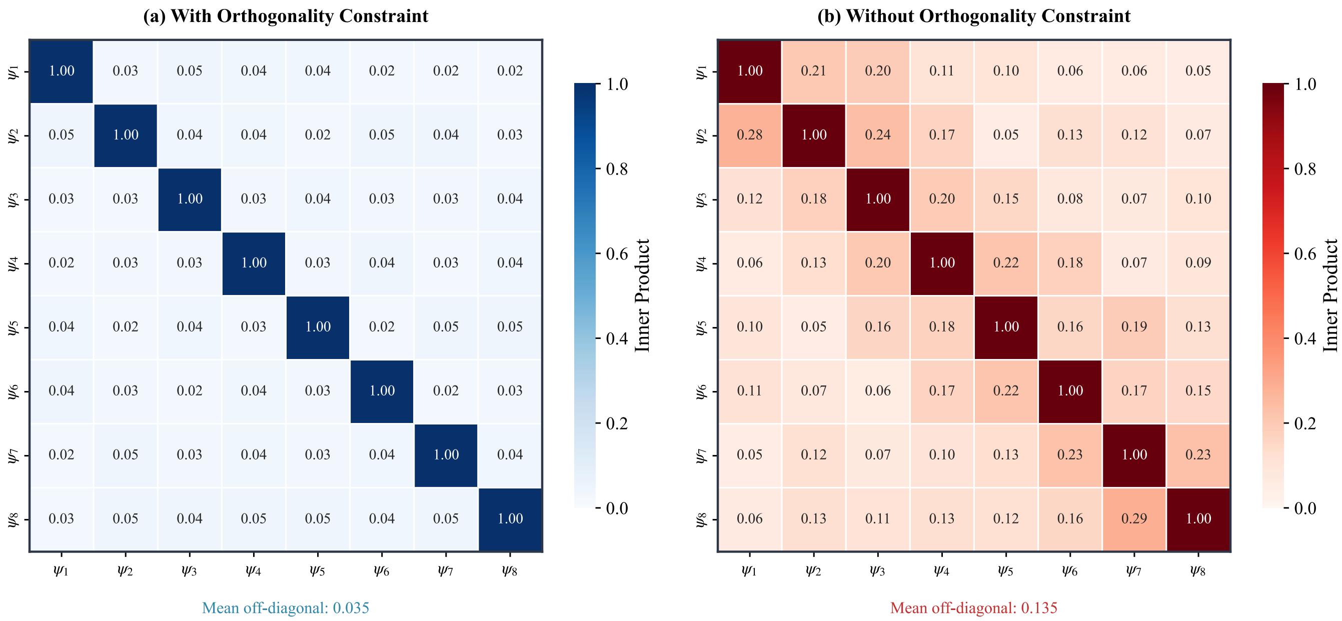

Figure 12. Orthogonality matrix between learned basis functions. Heatmap matrix rows and columns correspond to 8 adaptive basis functions. Matrix elements are inner product absolute values between two basis functions. Diagonal elements are basis function norms. Left side shows results with modal orthogonality constraint and right side shows results without constraint.

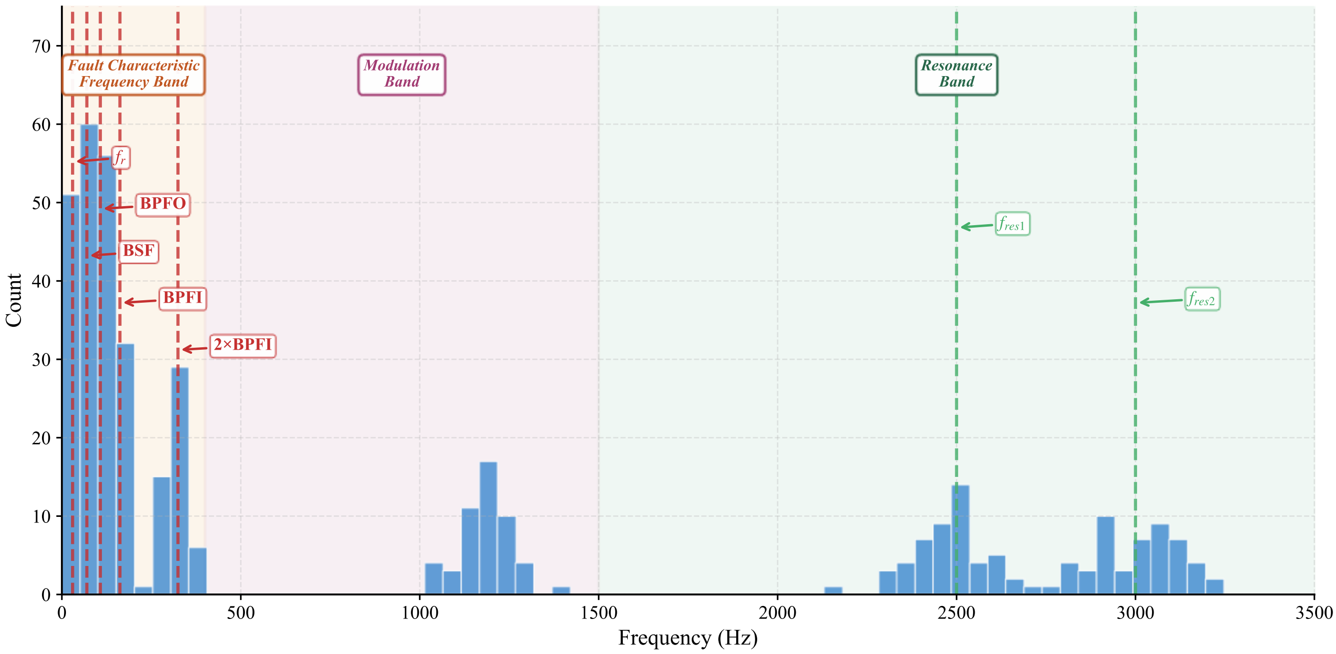

Figure 13. Histogram of adaptive basis function center frequency distribution and correspondence with theoretical characteristic frequencies. The horizontal axis is frequency in Hz and vertical axis is basis function occurrence count. Red dashed lines mark theoretical bearing characteristic frequency positions including =29.95Hz for rotation frequency, BSF=69.8Hz for ball fault, BPFO=107.4Hz for outer race fault, BPFI=162.2Hz for inner race fault, and 2×BPFI=324.4Hz for second harmonic. Green dashed lines mark resonance frequencies at =2500Hz and =3000Hz. Background colors distinguish fault characteristic frequency band (0-400Hz, orange), modulation band (400-1500Hz, purple), and resonance band (1500-3500Hz, green).

6.10. Parameter sensitivity analysis

The proposed method involves multiple hyperparameters. As shown in Figure 14, when the number of basis functions K increases from 2 to 8, accuracy shows upward trend, indicating that more basis functions can better represent multiple frequency components. When K exceeds 8, accuracy stabilizes and slightly decreases, possibly because excessive basis functions introduce redundant components. K=8 is selected as default setting.

Figure 15 shows influence of bidirectional cross-attention layer number on diagnostic performance. Single-layer cross-attention shows relatively low performance, indicating that shallow feature interaction is insufficient for full fusion of time-frequency and hypergraph representations. When the layer number increases to 2–3, accuracy improves. When the layer number exceeds 4, accuracy no longer improves, possibly due to overfitting. L=3 achieves optimal performance considering accuracy and model complexity.

The sensitivity of diagnostic performance to the causal threshold parameter γ is also examined. This parameter controls the density of the constructed hypergraph topology through the adaptive threshold $\theta_{cause}(t) = \mu_A(t) + \gamma\sigma_A(t)$ where $\mu_A(t)$ and $\sigma_A(t)$ denote the mean and standard deviation of the causal strength distribution at time t. When γ is set too small, the threshold falls below the mean causal strength, admitting a large number of weak causal connections into the hyperedge structure and introducing noise into the relational modeling. When γ is set too large, the threshold becomes excessively restrictive and the hypergraph degenerates toward isolated nodes, which undermines the high-order interaction modeling capability that motivates the hypergraph formulation. The default value $\gamma = 1.0$ retains connections that exceed one standard deviation above the mean causal strength, providing a balance between structural richness and noise suppression.

To quantitatively characterize the effect of γ, experiments are conducted with γ varying over {0.5, 0.75, 1.0, 1.25, 1.5, 2.0} while all other hyperparameters are held at their default values. Each configuration is repeated five times. The results are summarized in Table 1.

Table 1. Diagnostic accuracy (%) of the proposed method under different causal threshold values γ on the CWRU dataset. Results are reported as mean ± standard deviation over five repeated experiments.

γ	Hypergraph Topology	Accuracy (%)
0.5	Dense (many weak connections admitted)	96.54 ± 0.38
0.75	Moderately dense	97.82 ± 0.29
1.0	**Balanced (default)**	**98.73 ± 0.21**
1.25	Moderately sparse	98.35 ± 0.25
1.5	Sparse	97.61 ± 0.31
2.0	Highly sparse (isolated nodes appear)	96.08 ± 0.44

The results confirm the analysis above: accuracy peaks at $\gamma = 1.0$ and decreases in both directions. When $\gamma = 0.5$, the hypergraph admits excessive weak causal connections, introducing relational noise that reduces diagnostic accuracy by 2.19 percentage points relative to the default. When $\gamma = 2.0$, the threshold is too restrictive and the hypergraph degenerates toward an under-connected structure, resulting in a 2.65 percentage point decrease. The proposed default of $\gamma = 1.0$ provides a robust operating point, with accuracy remaining above 98% for γ in [0.75, 1.25], indicating that the method is not overly sensitive to moderate variations of this parameter.

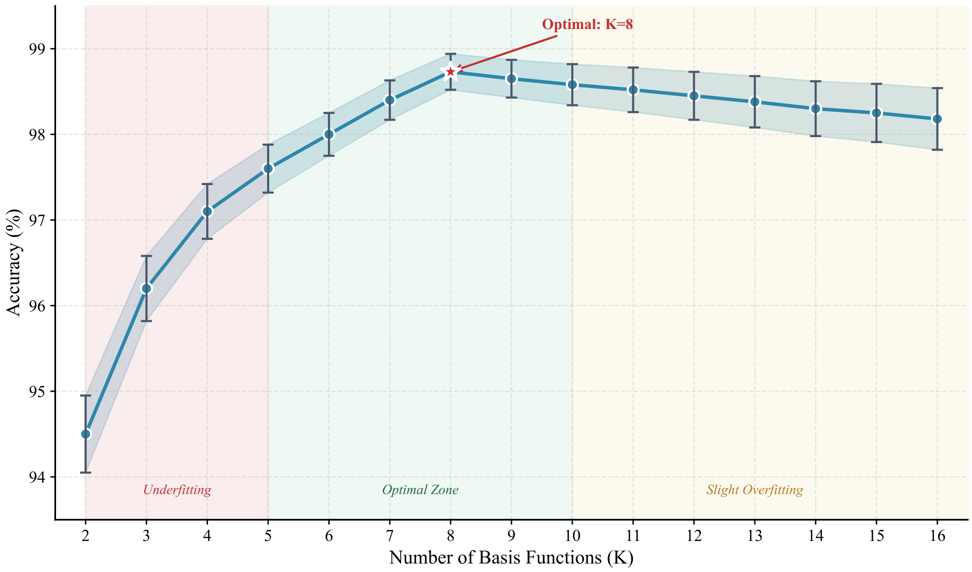

Figure 14. Influence curve of basis function number on diagnostic accuracy. The horizontal axis is basis function number K ranging from 2 to 16 and vertical axis is diagnostic accuracy. Error bars represent standard deviations from five repeated experiments.

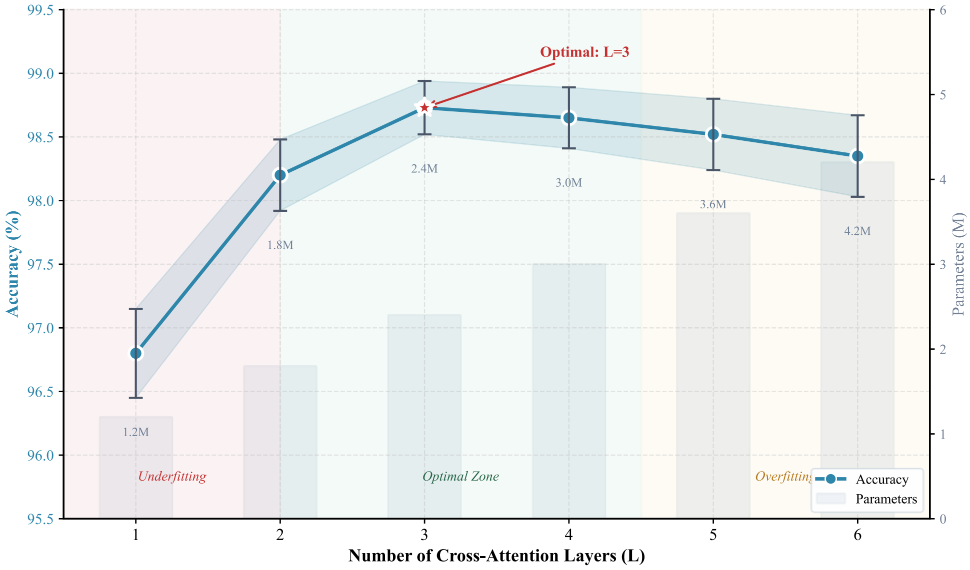

Figure 15. Influence curve of bidirectional cross-attention layer number on diagnostic accuracy. The horizontal axis is attention layer number L ranging from 1 to 6 and vertical axis is diagnostic accuracy. Model parameter counts are annotated for each layer number.

6.11. Computational complexity analysis

The practical applicability of the proposed method depends not only on diagnostic accuracy but also on its computational demands during training and inference. The transfer entropy computation constitutes the most computationally involved preprocessing step. For a system with N measurement points, the estimation of causal adjacency matrix entries requires evaluating N(N-1) directional transfer entropy values, each of which depends on conditional probability density estimation from finite-length time series. In the present experimental setting with N = 4 measurement points (DE-X, DE-Y, FE-X, FE-Y), this amounts to 12 pairwise computations per signal segment. For deployments involving a larger number of sensors, approximate entropy estimators or sensor-pair pre-screening strategies based on physical proximity or known mechanical coupling paths can be employed to reduce computational overhead.

The dynamic hypergraph update step runs once per forward pass and does not involve iterative optimization beyond standard neural network backpropagation, keeping inference-phase cost proportional to the hyperedge count rather than to the full pairwise causal computation. Pre-computing causal neighborhoods in an offline calibration phase and refreshing them at a lower frequency than the classification cycle represents a practical approach to reducing online inference latency in deployment contexts.

Table 2. Computational complexity comparison (NVIDIA RTX 3090, CWRU test set).

Method	Parameters (M)	Training Time (s/epoch) ↓	Inference Latency (ms) ↓
CNN [26]	**0.47**	**8.3**	**1.8**
WDCNN [25]	0.85	11.2	2.4
DRSN [27]	1.52	19.6	3.7
MRF-GCN [28]	1.83	26.4	5.1
TFT [29]	2.11	33.8	6.9
Proposed	2.40	47.2	9.6

As shown in Table 2, the proposed method carries a higher computational cost than simpler baseline methods, attributable primarily to the transfer entropy preprocessing and dynamic hypergraph update. The inference latency of 9.6 ms per sample is compatible with quasi-real-time monitoring applications. For stricter latency budgets, the offline causal neighborhood pre-computation strategy described above can further reduce online inference cost.

7. Conclusion and future work

This paper proposes a physical-causal guided adaptive time-frequency and hypergraph co-evolutionary method for industrial equipment monitoring. The framework integrates three collaborating modules. The first module learns parameterized time-frequency basis functions with energy conservation, frequency non-negativity, and modal orthogonality enforced through differentiable physical projection operators, ensuring that decomposed components conform to physical plausibility throughout training. The second module employs transfer entropy to discover inter-sensor causal relationships and converts them into dynamic hyperedge generation rules, constructing a hypergraph topology that captures high-order interaction patterns among measurement points. The third module applies bidirectional cross-attention to achieve iterative mutual enhancement between the time-frequency and hypergraph feature representations. The three modules are coupled through shared feature representations, so that the time-frequency decomposition informs the causal structure construction and the causal structure in turn guides the feature aggregation process. Experiments achieve 98.73% accuracy on CWRU, 91.77% cross-condition transfer, 82.56% at −6 dB SNR, and 82.4% cross-dataset generalization on unseen equipment.

The method is primarily suited to multi-sensor rotating machinery health monitoring scenarios where vibration signals exhibit non-stationary characteristics and multi-source coupling. Its design requires that measurement points be physically connected through a mechanical transmission path, so that transfer entropy can meaningfully reflect fault propagation directions, and that a sufficient length of signal history be available for reliable causal estimation. Under conditions of severely limited training data or very short observation windows, the reliability of the causal adjacency matrix estimation may be reduced. The cross-dataset experiments indicate that zero-shot transfer to equipment of a different mechanical type remains more challenging than within-type transfer, and domain adaptation strategies may offer further improvement in such settings.

Future work will pursue three directions. The first concerns the development of lightweight transfer entropy computation schemes suitable for online deployment, including approximate conditional probability estimators and sliding-window update strategies that reduce per-sample processing time to a level compatible with the latency constraints of edge hardware. The second concerns model compression and knowledge distillation approaches that reduce the parameter footprint of the bidirectional cross-attention module, making the full framework deployable on processors with constrained memory. The third concerns extending the framework from fault classification to remaining useful life estimation and degradation trend prediction, where the dynamic causal hypergraph structure may provide additional interpretable information about the stage of fault progression.

References

1. Hou B, Wang D, Xia T, Peng Z, Tsui KL. Difference mode decomposition for adaptive signal decomposition. *Mechanical Systems and Signal Processing* 2023; 191: 110203. https://doi.org/10.1016/j.ymssp.2023.110203.
2. Niu G, Liu E, Wang X, Ziehl P, Zhang B. Enhanced discriminate feature learning deep residual CNN for multitask bearing fault diagnosis with information fusion. *IEEE Transactions on Industrial Informatics* 2023; 19(1): 762-770. https://doi.org/10.1109/TII.2022.3165636.
3. Wang Z, Wu Z, Li X, Shao H, Han T, Xie M. Attention-aware temporal-spatial graph neural network with multi-sensor information fusion for fault diagnosis. *Knowledge-Based Systems* 2023; 278: 110891. https://doi.org/10.1016/j.knosys.2023.110891.
4. Wang H, Liu Z, Peng D, Zuo MJ. Interpretable convolutional neural network with multilayer wavelet for noise-robust machinery fault diagnosis. *Mechanical Systems and Signal Processing* 2023; 195: 110314. https://doi.org/10.1016/j.ymssp.2023.110314.
5. Wang F, Pena-Pena K, Qian W, Arce GR. T-HyperGNNs: Hypergraph neural networks via tensor representations. *IEEE Transactions on Neural Networks and Learning Systems* 2025; 36(3): 5044-5058. https://doi.org/10.1109/TNNLS.2024.3371382.
6. Gao Y, Feng Y, Ji S, Ji R. HGNN+: General hypergraph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2023; 45(3): 3181-3199. https://doi.org/10.1109/TPAMI.2022.3182052.
7. Miele S, Karve P, Mahadevan S. Multi-fidelity physics-informed machine learning for probabilistic damage diagnosis. *Reliability Engineering & System Safety* 2023; 235: 109243. https://doi.org/10.1016/j.ress.2023.109243.
8. Liu R, Zhang Q, Lin D, Zhang W, Ding SX. Causal disentangled graph neural network for fault diagnosis of complex industrial process. *IEEE Transactions on Industrial Informatics* 2025; 21(1): 386-395. https://doi.org/10.1109/TII.2024.3452246.
9. Miao Y, Zhang B, Li C, Lin J, Zhang D. Feature mode decomposition: New decomposition theory for rotating machinery fault diagnosis. *IEEE Transactions on Industrial Electronics* 2023; 70(2): 1949-1960. https://doi.org/10.1109/TIE.2022.3156156.
10. Chen Y, Mao Z, Hou X, Zhang Z, Zhang J, Jiang Z. Noise-robust adaptive feature mode decomposition method for accurate feature extraction in rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing* 2024; 211: 111213. https://doi.org/10.1016/j.ymssp.2024.111213.
11. Li M, Liu Y, Wang T, Chu F, Peng Z. Adaptive synchronous demodulation transform with application to analyzing multicomponent signals for machinery fault diagnostics. *Mechanical Systems and Signal Processing* 2023; 191: 110208. https://doi.org/10.1016/j.ymssp.2023.110208.
12. Chen Y, Zhang J, Zhao N, Mao Z, Jiang Z. Impact time domain decomposition: An adaptive decomposition method for multi-source impact signals based on envelope energy gradient characteristics. *Mechanical Systems and Signal Processing* 2024; 219: 111637. https://doi.org/10.1016/j.ymssp.2024.111637.
13. Yan X, Liu Y, Zhang C. Multiresolution hypergraph neural network for intelligent fault diagnosis. *IEEE Transactions on Instrumentation and Measurement* 2022; 71: 3525910. https://doi.org/10.1109/TIM.2022.3212532.
14. Zhao X, Zhu X, Liu J, Hu Y, Gao T, Zhao L, Yao J, Liu Z. Model-assisted multi-source fusion hypergraph convolutional neural networks for intelligent few-shot fault diagnosis to electro-hydrostatic actuator. *Information Fusion* 2024; 104: 102186. https://doi.org/10.1016/j.inffus.2023.102186.
15. Chen D, Liu R, Hu Q, Ding SX. Interaction-aware graph neural networks for fault diagnosis of complex industrial processes. *IEEE*

Transactions on Neural Networks and Learning Systems 2023; 34(9): 6015-6028. https://doi.org/10.1109/TNNLS.2021.3132376.
16. Wang S, Jiao X, Jing B, Pan J, Meng X, Huang Y, Pei S. Coupling fault diagnosis of bearings based on hypergraph neural network. *Sensors* 2024; 24(19): 6391. https://doi.org/10.3390/s24196391.
17. Qin D, Peng Z, Wu L. FCHG: Fuzzy cognitive hypergraph for interpretable fault detection. *Expert Systems with Applications* 2024; 255: 124700. https://doi.org/10.1016/j.eswa.2024.124700.
18. Ni Q, Ji JC, Halkon B, Feng K, Nandi AK. Physics-informed residual network (PIResNet) for rolling element bearing fault diagnostics. *Mechanical Systems and Signal Processing* 2023; 200: 110544. https://doi.org/10.1016/j.ymssp.2023.110544.
19. Xu Z, Zhao K, Wang J, Bashir M. Physics-informed probabilistic deep network with interpretable mechanism for trustworthy mechanical fault diagnosis. *Advanced Engineering Informatics* 2024; 62: 102806. https://doi.org/10.1016/j.aei.2024.102806.
20. Liu R, Zhang Q, Lin D, Zhang W, Ding SX. Causal intervention graph neural network for fault diagnosis of complex industrial processes. *Reliability Engineering & System Safety* 2024; 251: 110328. https://doi.org/10.1016/j.ress.2024.110328.
21. Zhang Z, Wu L. Graph neural network-based bearing fault diagnosis using Granger causality test. *Expert Systems with Applications* 2024; 242: 122827. https://doi.org/10.1016/j.eswa.2023.122827.
22. Wang H, Liu R, Ding SX, Hu Q, Li Z, Zhou H. Causal-trivial attention graph neural network for fault diagnosis of complex industrial processes. *IEEE Transactions on Industrial Informatics* 2024; 20(2): 1987-1996. https://doi.org/10.1109/TII.2023.3282979.
23. Smith WA, Randall RB. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mechanical Systems and Signal Processing* 2015; 64-65: 100-131. https://doi.org/10.1016/j.ymssp.2015.04.021.
24. Zhang M, Jiang Z, Feng K. Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump. *Mechanical Systems and Signal Processing* 2017; 93: 460-493. https://doi.org/10.1016/j.ymssp.2017.02.013.
25. Zhang W, Peng G, Li C, Chen Y, Zhang Z. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* 2017; 17(2): 425. https://doi.org/10.3390/s17020425.
26. Wen L, Li X, Gao L, Zhang Y. A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics* 2018; 65(7): 5990-5998. https://doi.org/10.1109/TIE.2017.2774777.
27. Zhao M, Zhong S, Fu X, Tang B, Pecht M. Deep residual shrinkage networks for fault diagnosis. *IEEE Transactions on Industrial Informatics* 2020; 16(7): 4681-4690. https://doi.org/10.1109/TII.2019.2943898.
28. Li T, Zhao Z, Sun C, Yan R, Chen X. Multireceptive field graph convolutional networks for machine fault diagnosis. *IEEE Transactions on Industrial Electronics* 2021; 68(12): 12739-12749. https://doi.org/10.1109/TIE.2020.3040669.
29. Ding Y, Jia M, Miao Q, Cao Y. A novel time-frequency Transformer based on self-attention mechanism and its application in fault diagnosis of rolling bearings. *Mechanical Systems and Signal Processing* 2022; 168: 108616. https://doi.org/10.1016/j.ymssp.2021.108616.
30. Li X, Zhang W, Ding Q. Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. *Signal Processing* 2019; 161: 136-154. https://doi.org/10.1016/j.sigpro.2019.03.019.
31. Nectoux P, Gouriveau R, Medjaher K, Ramasso E, Chebel-Morello B, Zerhouni N, Varnier C. PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In: *Proceedings of the IEEE International Conference on Prognostics and Health Management*; Denver, CO, USA. IEEE; 2012. p. 1-8. https://doi.org/10.1109/ICPHM.2012.6299167.
32. Bechhoefer E. Condition based maintenance fault database for testing diagnostics and prognostic algorithms. Society for Machinery Failure Prevention Technology (MFPT); 2012 [accessed 2024]. Available from: https://www.mfpt.org/fault-data-sets/.
33. Lessmeier C, Kimotho JK, Zimmer D, Sextro W. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. *PHM Society European Conference* 2016; 3(1). https://doi.org/10.36001/phme.2016.v3i1.1577.
34. Shao S, McAleer S, Yan R, Baldi P. Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Transactions on Industrial Informatics* 2019; 15(4): 2446-2455. https://doi.org/10.1109/TII.2018.2864759.