ICJK
Chinese Journal of Information Fusion
http://dx.doi.org/10.62762/CJIF.2026.701879

RESEARCH ARTICLE

An Evidence Fusion Approach for Rolling Bearing Fault Diagnosis Using PredVAR-Extracted Latent Variables

Zhongxi Zhu[1], **Weijun Lv**[1] and **Yanfang Mo**[1,*]

[1] School of Data Science, Lingnan University, Hong Kong, China

Abstract

Conventional data-driven fault diagnosis for rolling-element bearings often uses static dimensionality-reduction methods, such as principal component analysis (PCA). Such methods may preserve load-related static variation and produce inconsistent evidence when several classifiers are fused. This paper proposes a dynamic feature extraction and evidence fusion method based on the probabilistic reduced-dimensional vector autoregressive (PredVAR) model proposed in [Mo and Qin, 2025, Automatica]. The raw vibration signal is first reconstructed by delay embedding, and PredVAR is then used to extract predictable dynamic latent variables (DLVs). The extracted DLVs are converted into compact dynamic feature vectors for learning vector quantization (LVQ) and decision tree (DT) classifiers, whose outputs are combined by Dempster-Shafer (D-S) evidence theory. Experiments on a mixed-load CWRU bearing dataset show that the proposed PredVAR + D-S framework achieves an average accuracy of 99.91% with a standard deviation of 0.0014 under the main 20-run protocol. The D-S conflict coefficient is reduced from 0.0195 to 0.0056 compared with PCA-based features, indicating more consistent classifier evidence. The proposed method can be positioned as a compact and interpretable dynamic-fusion approach. In a limited-label setting with 50 labeled windows per class, PredVAR + D-S obtains the highest mean accuracy among the compared methods, suggesting an advantage when labeled fault data are scarce.

Keywords: fault diagnosis, dynamic latent variable, PredVAR, Dempster-Shafer evidence theory, information fusion.

1 Introduction

Rolling-element bearings are key components in rotating machinery, and their condition has a direct influence on the safety and continuity of equipment operation. With the increasing availability of vibration measurements, data-driven fault diagnosis has become a common way to identify bearing health states from sensor data [1–3]. A typical diagnostic pipeline contains signal acquisition, feature extraction, and fault classification. Among these steps, feature extraction is especially important because it determines what information is passed to the classifier or the fusion layer.

Academic Editor:
Fuyuan Xiao

Submitted: 17 January 2026
Revised: 21 May 2026
Accepted: 03 July 2026
Published: 17 Nov 2026

Vol. 3, **No.** 4, 2026.
10.62762/CJIF.2026.701879

***Corresponding author:**
✉ Yanfang Mo
yanfangmo@LN.edu.hk

Citation
Zhu, Z., Lv, W., & Mo, Y. (2026). An Evidence Fusion Approach for Rolling Bearing Fault Diagnosis Using PredVAR-Extracted Latent Variables. *Chinese Journal of Information Fusion*, 3(4), 238–250.

© 2026 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Many traditional methods describe vibration windows by statistical indicators in the time or frequency domain, such as root mean square (RMS), kurtosis, and spectral energy [4, 5]. Since these handcrafted features can be redundant, static dimensionality reduction methods, especially principal component analysis (PCA), are often used before classification [6, 7], and PCA has also been combined with deep networks in other fault diagnosis tasks such as wind turbines [8]. For example, PCA-based feature pipelines have been combined with Dempster-Shafer (D-S) evidence fusion [9], while learning vector quantization (LVQ) networks [10, 11] and decision trees (DTs) [12, 13] are common lightweight base classifiers. These methods are simple and efficient, but they usually treat each vibration window as an independent static sample. As a result, the temporal evolution inside the raw signal is compressed into overall distributional quantities, and part of the dynamic information related to fault impacts may be weakened.

Vibration signals from rotating machinery are generated by dynamic systems and often show non-stationary behavior under variable operating conditions [14–16]. Bearing faults may appear as repeated impulses, modulation effects, or other predictable patterns related to the system motion. They are therefore not only changes in static statistical distributions. For this reason, feature extraction should not only preserve high variance, but should also focus on predictable dynamics. The probabilistic reduced-dimensional vector autoregressive (PredVAR) model [17] is suitable for this purpose. It identifies dynamic latent variables (DLVs) through a reduced-dimensional VAR model with oblique projections. Related dynamic dimensionality-reduction ideas, such as dynamic PCA, have also been developed for dynamic data modeling and process monitoring [18]. By applying PredVAR to delay-embedded vibration sequences, with the phase-space reconstruction idea commonly used in bearing signal analysis [19, 20], the present work aims to extract DLVs that retain predictable fault-related dynamics while suppressing static noise components.

In a multi-classifier fusion system, the quality of the feature representation also affects the reliability of the fused decision. Studies on evidence theory, including complex evidence theory, conflict handling in belief-function combination, and basic belief assignment determination, have shown that the handling of uncertain or conflicting evidence remains important in multi-source fusion [21–23]. D-S evidence theory has also been used in machinery and bearing diagnosis tasks [24, 25]. When two base classifiers are driven by noisy static features, their posterior outputs may point to different classes, and the conflict coefficient in D-S fusion increases. This paper therefore combines PredVAR-based dynamic feature extraction with D-S evidence fusion. The key idea is not only to replace PCA with PredVAR, but also to improve the source of evidence before D-S fusion. By supplying LVQ and DT classifiers with predictable DLV-based features, the proposed framework reduces evidence conflict and makes the final fused decision more stable.

The experimental validation is conducted on the CWRU bearing dataset [26]. Residual networks have been used to improve bearing fault feature extraction under noisy labels and noisy environments [27], and physics-informed residual networks have been developed for rolling bearing fault diagnostics [28]. Transformer-based models combined with residual networks have also been introduced to capture longer-range temporal or spectral dependencies [29]. In addition, transfer-learning strategies have been studied to reduce the influence of load changes and limited labeled samples [30]. These methods provide strong classification ability, but they often require sufficient labeled data, careful training, and a relatively large number of model parameters. Their learned features are also less direct to interpret in terms of the underlying machine dynamics.

The main contributions of this work are summarized as follows:

1. A PredVAR-based dynamic feature extraction pipeline is developed for bearing fault diagnosis. Raw vibration windows are first reconstructed by delay embedding, and PredVAR is then used to extract predictable DLVs. Unlike PCA, which selects directions of large static variance, PredVAR selects a compact dynamic subspace with stronger one-step predictability.

2. A PredVAR-driven D-S evidence fusion framework is constructed. LVQ and DT are used as two complementary base classifiers, and their basic probability assignments (BPAs) are fused by D-S evidence theory. The D-S conflict coefficient K is used to examine the consistency of the generated evidence, showing that PredVAR improves the evidence source before the final D-S fusion step.

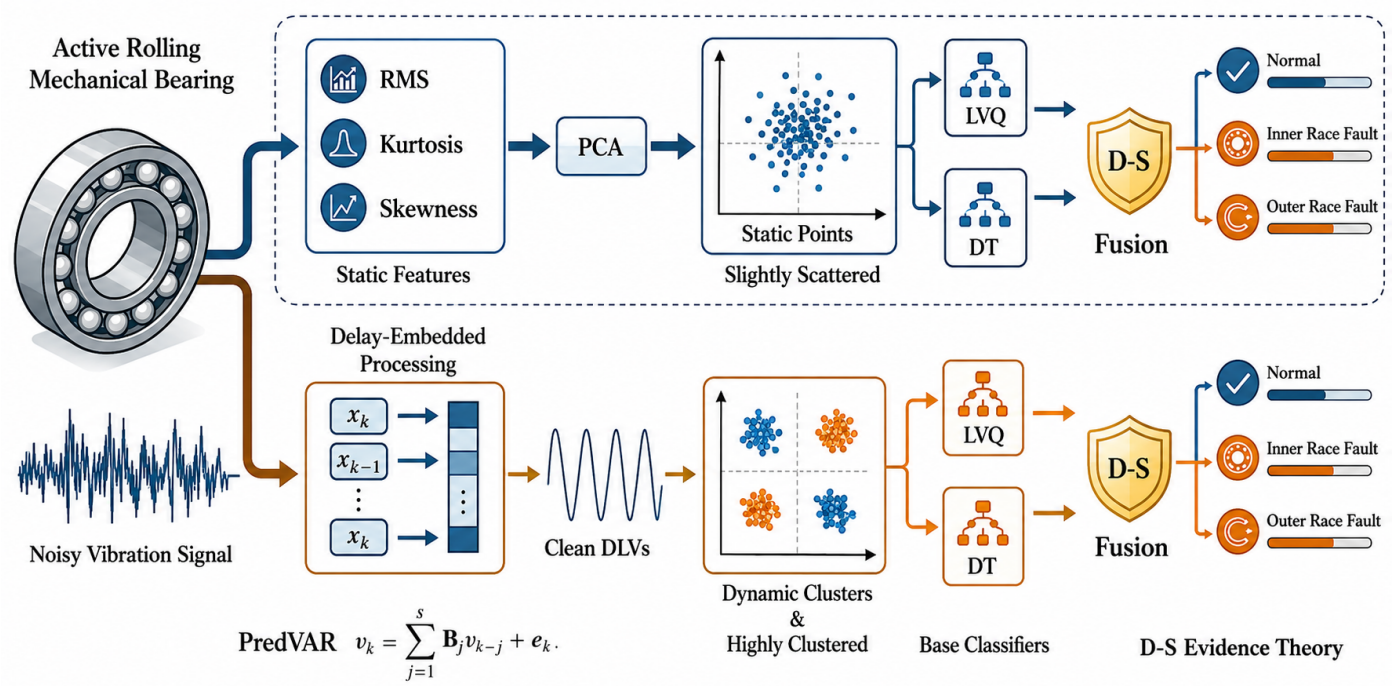


Figure 1. The overall architecture of the proposed PredVAR-driven information fusion framework.

3. The method is evaluated under the mixed-load CWRU protocol through accuracy, stability, evidence-conflict, sensitivity, computational-cost, and label-efficiency analyses. Recent neural baselines are also considered to position the method against high-capacity classifiers. The full-data comparison shows that the 1D-Transformer and the ResNet-based transfer-learning baseline can be stronger in pure accuracy under the same full-data random-window protocol. However, the proposed method provides a compact and interpretable dynamic-fusion route with low evidence conflict and a clear advantage in the limited-label setting considered here.

2 Methodology

In this section, we elaborate on the proposed information fusion framework based on PredVAR feature extraction. Unlike traditional static methods (e.g., PCA) that maximize variance, the proposed architecture employs the PredVAR model to maximize predictability, thereby capturing the intrinsic temporal evolution of machinery fault signals. The overall architecture of the proposed diagnostic system is illustrated in Figure 1.

2.1 PredVAR Model Formulation

To lay the mathematical foundation for our dynamic feature extraction, we first revisit the PredVAR model introduced by Mo and Qin [17]. Consider the high-dimensional measurement time series sequence $\{\boldsymbol{y}_k \in \Re^p\}_{k=1}^N$, where p denotes the dimension of the measurement space (or the embedded phase space dimension). We assume $\{\boldsymbol{y}_k\}$ is zero-mean stationary. The primary objective of PredVAR is to decompose the measurement space into a *dynamic subspace* governed by predictable variations and a *static noise subspace* dominated by unpredictable noise.

The PredVAR model is defined by the following pair of equations:

$$\boldsymbol{y}_k = \mathbf{P}\boldsymbol{v}_k + \bar{\mathbf{P}}\bar{\boldsymbol{\varepsilon}}_k, \quad \bar{\boldsymbol{\varepsilon}}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\bar{\boldsymbol{\varepsilon}}}), \tag{1}$$

$$\boldsymbol{v}_k = \sum\nolimits_{j=1}^{s} \mathbf{B}_j \boldsymbol{v}_{k-j} + \boldsymbol{\varepsilon}_k, \quad \boldsymbol{\varepsilon}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}), \tag{2}$$

where

1. $\boldsymbol{v}_k \in \Re^\ell$ represents the DLVs, whose dimension is smaller than the measurement-space dimension, i.e., $\ell < p$. These variables capture the essential temporal dynamics of the system.

2. $\mathbf{P} \in \Re^{p\times\ell}$ is the loading matrix for the dynamic subspace. $\bar{\mathbf{P}} \in \Re^{p\times(p-\ell)}$ corresponds to the static noise subspace.

3. $\bar{\boldsymbol{\varepsilon}}_k \in \Re^{p-\ell}$ is the static noise vector, which is serially uncorrelated.

4. Equation (2) describes the vector autoregressive (VAR) dynamics of the latent variables with VAR

order s, where $\mathbf{B}_j \in \Re^{\ell\times\ell}$ are the corresponding VAR coefficient matrices in the DLV subspace.

5. $\boldsymbol{\varepsilon}_k \in \Re^{\ell}$ is the innovation of the DLVs.

Distinct from static dimensionality reduction, the DLVs $\boldsymbol{v}_k$ in this model are full-dimensional dynamic, meaning they contain no nonzero purely static linear combinations. This property helps ensure that the extracted features represent predictable dynamic information (e.g., fault-induced periodic impulses) rather than static background energy.

2.2 Dynamic Feature Extraction via Oblique Projection

To explicitly extract the dynamic features $\boldsymbol{v}_k$ from the observed data $\boldsymbol{y}_k$, we introduce the corresponding weighting matrices $\mathbf{R} \in \Re^{p\times\ell}$ and $\bar{\mathbf{R}} \in \Re^{p\times(p-\ell)}$ that satisfy the following identity:

$$\begin{bmatrix}\mathbf{R} & \bar{\mathbf{R}}\end{bmatrix}^{\intercal}\begin{bmatrix}\mathbf{P} & \bar{\mathbf{P}}\end{bmatrix} = \mathbf{I}. \tag{3}$$

Specifically, this implies $\mathbf{R}^{\intercal}\mathbf{P} = \mathbf{I}$ and $\mathbf{R}^{\intercal}\bar{\mathbf{P}} = \mathbf{0}$. By pre-multiplying Eq. (1) by $\mathbf{R}^{\intercal}$, we obtain the explicit realization of the DLVs:

$$\boldsymbol{v}_k = \mathbf{R}^{\intercal}\boldsymbol{y}_k. \tag{4}$$

Equation (4) serves as the core feature extraction mechanism in our framework. The matrix $\mathbf{R}$ functions as a dynamic filter that projects the high-dimensional observation $\boldsymbol{y}_k$ onto the low-dimensional dynamic manifold, filtering out the static noise $\bar{\boldsymbol{\varepsilon}}_k$.

The optimal one-step ahead prediction of the DLVs is given by:

$$\hat{\boldsymbol{v}}_k = \sum\nolimits_{j=1}^{s} \mathbf{B}_j \boldsymbol{v}_{k-j}. \tag{5}$$

Consequently, the measurement can be decomposed into a predictable component and a generalized innovation term:

$$\boldsymbol{y}_k = \mathbf{P}\hat{\boldsymbol{v}}_k + \boldsymbol{e}_k, \tag{6}$$

where $\boldsymbol{e}_k$ represents the innovation of $\boldsymbol{y}_k$.

To quantitatively evaluate the modeling efficacy and the predictability of the extracted DLVs upon convergence, the Final Prediction Error (FPE) criterion is utilized. The FPE metric evaluates the generalized variance of the latent prediction errors:

$$\mathrm{FPE} = \det(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}})\left(\frac{N+s\ell}{N-s\ell}\right)^{\ell}, \tag{7}$$

where $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}$ is the estimated covariance matrix of the latent innovations (prediction errors), N is the number of samples, ℓ is the dimension of the DLVs, and s is the VAR order. The multiplicative term acts as a penalty factor for model complexity to prevent overfitting.

In the normalized phase space, the FPE metric is used as a relative predictability indicator. An FPE ≈ 0 signifies that the generalized innovation variance is negligible compared to the normalized raw-data variance, indicating that the extracted DLVs exhibit highly predictable temporal dynamics. Conversely, an FPE ≈ 1.0 indicates limited predictive power, implying the input data is mainly dominated by unpredictable static noise lacking temporal causality. In our framework, this metric serves as a theoretical indicator to validate the dynamic nature of the extracted features.

2.3 Optimal Parameter Estimation

The objective of model identification is to estimate the parameters $\{\mathbf{B}_j\}_{j=1}^{s}$, $\mathbf{P}$, and $\mathbf{R}$ such that the predictability of the extracted DLVs is maximized. This is achieved by minimizing the determinant of the innovation covariance matrix $\det(\boldsymbol{\Sigma}_{\boldsymbol{e}})$, i.e., the generalized prediction error variance.

Since the parameters are mutually coupled, an iterative solution based on the Expectation-Maximization (EM) framework and Instrumental Variables (IV) is employed. The optimization procedure involves the following key steps:

2.3.1 *Data Normalization*

First, the data are whitened to remove static correlations. Let $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{y}} = \mathbf{U}\mathbf{D}\mathbf{U}^{\intercal}$ be the eigenvalue decomposition of the sample covariance. The normalized vector is defined as $\boldsymbol{y}_k^* = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}^{\intercal}\boldsymbol{y}_k$.

The problem is then converted to finding the normalized loading matrix $\mathbf{P}^*$.

2.3.2 *Iterative Refinement*

The solution is obtained by iteratively refining the subspace estimate. In each iteration:

1. E-Step (Prediction): Given the current estimate of the subspace $\hat{\mathbf{P}}^*$, the DLVs are estimated, and the VAR coefficients $\hat{\mathbf{B}}_j$ are updated via least squares regression on the latent sequence. The one-step predictions $\hat{\mathbf{V}}_s$ are then computed.

2. M-Step (Subspace Update): The normalized loading matrix $\mathbf{P}^*$ is updated by solving the following eigenvalue problem at each refinement

step:

$$\frac{1}{N}\mathbf{Y}_s^{*\intercal}\mathbf{\Pi}_{\hat{\mathbf{V}}_s}\mathbf{Y}_s^{*}\mathbf{q}=\lambda\mathbf{q}, \tag{8}$$

where $\mathbf{Y}_s^*$ is the Hankel matrix of normalized measurements in the embedded training sequence, and $\mathbf{\Pi}_{\hat{\mathbf{V}}_s}=\hat{\mathbf{V}}_s(\hat{\mathbf{V}}_s^{\intercal}\hat{\mathbf{V}}_s)^{-1}\hat{\mathbf{V}}_s^{\intercal}$ is the projection matrix onto the space spanned by the predicted DLVs at the current iteration.

The matrix $\hat{\mathbf{P}}^*$ consists of the eigenvectors corresponding to the ℓ largest eigenvalues of Eq. (8). This process effectively aligns the subspace $\mathbf{P}$ with the direction of maximum predictability.

2.3.3 *Feature Recovery*

Upon convergence, the parameters are de-normalized to the original space:

$$\hat{\mathbf{P}}=\mathbf{U}\mathbf{D}^{\frac{1}{2}}\hat{\mathbf{P}}^{*},\quad \hat{\mathbf{R}}=\mathbf{U}\mathbf{D}^{-\frac{1}{2}}\hat{\mathbf{P}}^{*}. \tag{9}$$

The final dynamic features for fault diagnosis are obtained by projecting the delay-embedded test data using the estimated $\hat{\mathbf{R}}$ via Eq. (4).

To ensure that the extracted dynamic latent variables share a consistent physical coordinate system across different samples, the optimal parameters (including the projection matrix $\mathbf{R}$) are determined via a global fitting strategy during the offline training phase. Specifically, a subset of raw vibration sequences is uniformly sampled across all health states. After delay embedding, these matrices are concatenated into a common global observation space to fit a unified PredVAR model. The resulting projection matrix $\mathbf{R}$ then serves as a universal dynamic filter for all subsequent training and testing samples.

2.4 Hybrid Feature Vector Construction

To bridge the dynamic modeling layer and the downstream classification layer, the extracted dynamic sequences must be converted into fixed-length diagnostic feature vectors.

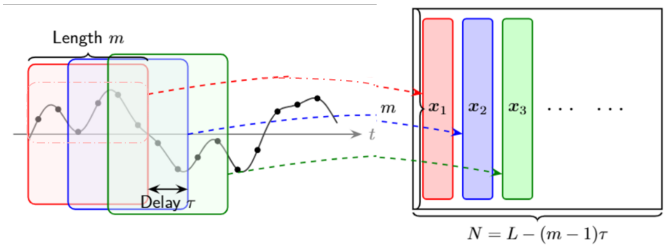


Figure 2. Delay embedding.

As illustrated in Figure 2, given a 1D raw discrete vibration signal window of length L, denoted as $x(1),x(2),\ldots,x(L)$, the initial phase space reconstruction via delay embedding maps sliding segments of the scalar time series into column vectors $\boldsymbol{x}_k$. For a given embedding dimension m and time delay τ, the k-th state vector is constructed as:

$$\boldsymbol{x}_k=[x(k),x(k+\tau),\ldots,x(k+(m-1)\tau)]^{\intercal}\in\Re^m. \tag{10}$$

By sliding the temporal window from $k=1$ to N, this operation converts the 1D signal into a multivariate trajectory matrix $\mathbf{X}\in\Re^{m\times N}$ with $N=L-(m-1)\tau$:

$$\mathbf{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N],\quad X_{i,k}=x\big(k+(i-1)\tau\big), \tag{11}$$

where $i=1,2,\ldots,m$ and $k=1,2,\ldots,N$. This compact notation is equivalent to the standard delay-embedding matrix and keeps the expression readable.

With the phase space reconstructed, the m-dimensional state vector $\boldsymbol{x}_k$ serves as the direct mathematical input to the PredVAR dynamic filter. Specifically, by setting the measurement space dimension $p=m$, we formally define the observation vector as $\boldsymbol{y}_k\triangleq\boldsymbol{x}_k$. The PredVAR algorithm subsequently processes this multivariate sequence $\{\boldsymbol{y}_k\}_{k=1}^N$ to extract a corresponding sequence of low-dimensional dynamic latent variables, $\{\boldsymbol{v}_k\in\Re^{\ell}\}_{k=1}^N$.

To evaluate the temporal dynamics of each latent dimension independently over the entire window, we rearrange this sequence. Let $v_k^{(i)}$ denote the i-th component of the DLV vector at time step k. The complete time-series sequence for the i-th DLV is thus defined as an N-dimensional column vector:

$$\boldsymbol{v}^{(i)}=\left[v_1^{(i)},v_2^{(i)},\ldots,v_N^{(i)}\right]^{\intercal}\in\Re^N,\quad i=1,2,\ldots,\ell. \tag{12}$$

Traditional feature extraction methods often compute statistical indicators directly from raw, noise-corrupted signals. In contrast, the DLV sequences $\boldsymbol{v}^{(i)}$ in our framework have already been filtered by the PredVAR dynamic filter. Because PredVAR explicitly projects the delay-embedded phase space onto a predictable dynamic subspace, the resulting $\boldsymbol{v}^{(i)}$ are expected to emphasize predictable, fault-related periodic impacts and to suppress the static noise space.

Specifically, to fulfill the fixed-length input requirements of the downstream classifiers, these dynamically filtered sequences undergo a statistical pooling process. For each latent channel $\boldsymbol{v}^{(i)}$, three representative temporal metrics are calculated: root

mean square (RMS) to quantify the deterministic dynamic energy, kurtosis to measure the impact impulsiveness, and peak-to-peak (PTP) value to capture the maximum amplitude span of the localized fault impacts. By concatenating these three metrics across all ℓ latent variables, a compact hybrid dynamic feature vector $\boldsymbol{x}_{dyn} \in \Re^{3\ell}$ is constructed. This gives the base classifiers compact dynamic evidence while preserving the physical meaning of the extracted features.

2.5 Base Classification Layer

D-S decision-level fusion benefits from evidence sources that are both reliable and complementary. If the base classifiers have very similar decision mechanisms, they may make highly correlated errors, and the fusion result may not necessarily improve the final diagnosis. For this reason, the proposed framework uses two different classifiers: LVQ, which is distance-based, and DT, which is rule-based.

The LVQ classifier evaluates the global structure of the dynamic feature space $\boldsymbol{x}_{dyn} \in \Re^{3\ell}$. It maintains a set of class-related prototype vectors $\mathcal{W} = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_M\}$. For a given input, LVQ assigns the class label of the winning prototype $\boldsymbol{w}_c$ by minimizing the Euclidean distance:

$$c = \arg \min_{j \in \{1,\ldots,M\}} \|\boldsymbol{x}_{dyn} - \boldsymbol{w}_j\|_2. \tag{13}$$

This classifier forms distance-based decision regions and is sensitive to the relative locations of samples in the feature space.

The DT classifier provides a different decision mechanism. It recursively partitions the feature space with axis-aligned splitting rules. At each internal node, it selects a feature f from $\boldsymbol{x}_{dyn}$ that maximizes the Information Gain (IG):

$$IG(D, f) = Ent(D) - \sum_{v=1}^{V} \frac{|D^v|}{|D|} Ent(D^v), \tag{14}$$

where $Ent(D)$ denotes the Shannon entropy of the current dataset D, and D^v denotes the subset produced by the v-th split of feature f. Compared with LVQ, DT is more sensitive to local feature thresholds and rule-based partitions.

The two classifiers are therefore not selected for their individual complexity, but because their decision mechanisms are deliberately different. This difference encourages complementary diagnostic evidence and reduces the chance that the fusion layer receives two nearly identical error patterns.

2.6 Decision-Level Fusion via D-S Evidence Theory

D-S evidence theory is used to combine the outputs of LVQ and DT.

Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_C\}$ be the frame of discernment containing all C bearing health states. In this work, Θ = {Normal, Inner Race Fault, Outer Race Fault}. For each testing sample, the classifier outputs are first converted into BPAs.

For DT, the probability $P_{DT}(\theta_i|\boldsymbol{x})$ is obtained from the class distribution at the corresponding reached leaf node. For LVQ, in our distance-weighted nearest-prototype implementation, the probability $P_{LVQ}(\theta_i|\boldsymbol{x})$ is computed from inverse-distance weights. The resulting posterior probabilities are then normalized to form the BPA for classifier j:

$$\mu_j(\{\theta_i\}) = \frac{P_j(\theta_i|\boldsymbol{x})}{\sum_{c=1}^{C} P_j(\theta_c|\boldsymbol{x}) + \delta}, \quad j \in \{1, 2\}, \tag{15}$$

where $\delta = 10^{-12}$ is used only for numerical stability. Since both BPAs are assigned only to singleton classes and no mass is placed on Θ, Dempster's rule reduces to a normalized product of the two class distributions, and K measures the disagreement between the two classifiers.

The BPAs are fused by Dempster's orthogonal sum rule. For any hypothesis $A \subseteq \Theta$, the fused mass is

$$\begin{aligned} \mu_{fused}(A) &= (\mu_1 \oplus \mu_2)(A) \\ &= \frac{1}{1-K} \sum_{B_1 \cap B_2 = A} \mu_1(B_1) \cdot \mu_2(B_2), \end{aligned} \tag{16}$$

where K is the conflict coefficient:

$$K = \sum_{B_1 \cap B_2 = \emptyset} \mu_1(B_1) \cdot \mu_2(B_2). \tag{17}$$

The final diagnosis is given by

$$\hat{\theta} = \arg \max_{\theta_i \in \Theta} \mu_{fused}(\{\theta_i\}). \tag{18}$$

The value of K is used as an evidence-conflict indicator in this framework. If the two classifiers assign large masses to different classes, K increases and the normalized fusion result may become less stable. In the proposed method, PredVAR is used before classification to provide a cleaner dynamic representation. The experimental conflict analysis in Section 3.6 shows that this representation reduces the average conflict coefficient compared with PCA-based static features.

2.7 Synergy Between PredVAR and D-S Fusion

The proposed framework is not simply a serial connection of a feature extractor and a fusion rule. The role of PredVAR is to change the quality of the evidence before D-S fusion is applied. PCA selects components by static variance, and those components may contain load-related amplitude variation or background noise. When LVQ and DT are trained on these variance-oriented features, their BPAs are more likely to concentrate on different fault classes, thereby increasing the conflict coefficient K.

PredVAR instead extracts DLVs by focusing on one-step predictability in the delay-embedded signal. Fault-related impulses and modulation patterns tend to be repeatable in time, while static noise is less predictable. Therefore, the DLV-based feature vector gives the base classifiers a more consistent description of the same fault state. In this sense, PredVAR and D-S fusion work together: PredVAR reduces the conflict in the generated evidence, and D-S fusion combines the remaining complementary information from LVQ and DT.

3 Experiments and Analysis

Experiments were conducted on the CWRU bearing dataset [26] to evaluate the proposed PredVAR-based feature extraction and evidence fusion framework. The main purpose of this section is to compare dynamic latent-variable features with static PCA features under the same fusion structure. Additional analyses are included to examine recent neural baselines, evidence conflict, hyperparameter sensitivity, computational cost, and label efficiency.

3.1 Dataset Description

The experimental data are from the Case Western Reserve University (CWRU) bearing data center [26]. A mixed-condition dataset was constructed by combining different motor loads and fault severities.

1. Data source: Drive-end (DE) accelerometer signals sampled at 12 kHz.
2. Health states: Three classes are considered: Normal Baseline, Inner Race Fault, and Outer Race Fault located at 6:00.
3. Mixed conditions: The dataset includes four motor loads, 0, 1, 2, and 3 hp, and three fault diameters, 0.007, 0.014, and 0.021 inches.

The raw signals were segmented into non-overlapping windows of length $L = 2048$. For the main comparison, the samples were randomly split into 70% training data and 30% testing data in each run. This split was repeated for 20 independent Monte Carlo runs. For all methods, the feature extractors, scalers, PCA model, PredVAR model, and classifiers were strictly fitted only on the training data within each run.

3.2 Experimental Setup

The experiments compare the proposed framework with both the PCA-based fusion baseline and representative external classifiers.

1. Core fusion comparison:

1. **PCA + D-S fusion:** Ten time-domain statistical features are first extracted from each raw signal window. PCA is then used to reduce the feature dimension to $\ell = 4$. The resulting principal components are fed into LVQ and DT classifiers, and the two outputs are fused by D-S evidence theory.
2. **PredVAR + D-S fusion:** The proposed method applies PredVAR to delay-embedded raw signals. Four DLVs are extracted, and RMS, kurtosis, and peak-to-peak (PTP) values are calculated for each latent channel. The resulting 3ℓ-dimensional feature vector is fed into the same LVQ and DT classifiers, followed by D-S fusion.

2. External baselines:

1. **Static + SVM:** Statistical features are classified by a Support Vector Machine with an RBF kernel.
2. **1D-CNN:** A convolutional neural network directly processes raw vibration windows.
3. **1D-ResNet:** A residual convolutional model is used as a stronger raw-signal neural baseline.
4. **1D-Transformer:** A temporal attention-based model is used to check the performance of sequence-level neural feature learning.
5. **ResNet-transfer:** A ResNet-based transfer-learning baseline is included to examine the benefit of pretrained or source-domain-initialized representations for cross-condition diagnosis.

All methods follow the same data-handling rules. In each Monte Carlo run, feature extractors, scalers, dimensionality-reduction models, PredVAR projections, and classifiers are fitted only on the training data. The recent neural baselines are discussed separately in Section 3.3.2 because they are

used mainly to position the proposed interpretable fusion framework against high-capacity data-driven models.

The default PredVAR parameters are set to $m = 64$, $\tau = 1$, VAR order $s = 2$, and latent dimension $\ell = 4$. A one-factor-at-a-time sensitivity study is added to examine the influence of these parameters.

3.3 Diagnosis Performance and Robustness Analysis

The diagnostic results across 20 random runs are summarized in Table 1. This table focuses on the comparison between the static PCA-based fusion structure and the proposed PredVAR-based fusion structure, with SVM and 1D-CNN included as reference baselines.

3.3.1 Fusion Efficacy and Variance Characteristics

The results in Table 1 show that replacing PCA with PredVAR improves both the individual base classifiers and the fused result. With PCA features, the D-S fusion accuracy is 97.20%, and the standard deviation is 0.0042. With PredVAR features, LVQ and DT both exceed 98%, and the fused PredVAR + D-S result reaches 99.91% with a standard deviation of 0.0014.

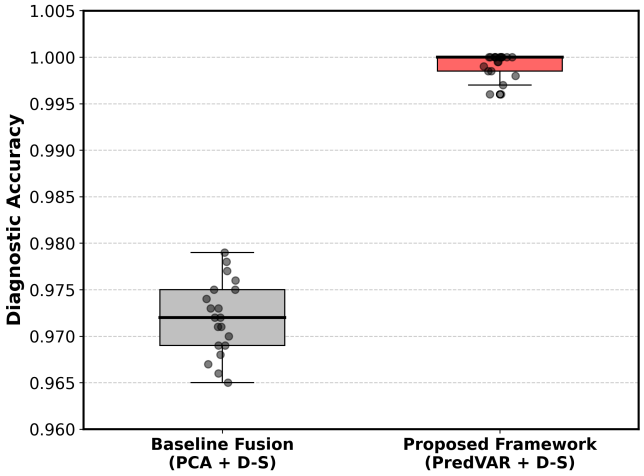


Figure 3. Boxplot of diagnosis accuracy.

The PCA-based result suggests that static principal components are not always suitable for evidence fusion. Since PCA preserves high-variance directions, it may keep load-related amplitude changes together with fault information. This can make LVQ and DT disagree on some samples, which is reflected by the larger variance of PCA + D-S. As shown in Figure 3, the accuracy distribution of PredVAR + D-S is more concentrated within the upper performance range across repeated runs, whereas the PCA-based fusion exhibits a wider spread. In contrast, PredVAR extracts predictable DLVs before classification. The resulting features provide more consistent evidence to the two base classifiers, which is consistent with the lower variance of the fused decision.

The internal ablation results in Figure 4 show that D-S fusion improves the two PredVAR-driven base classifiers under the same dynamic-feature representation. The run-by-run trajectories in Figure 5

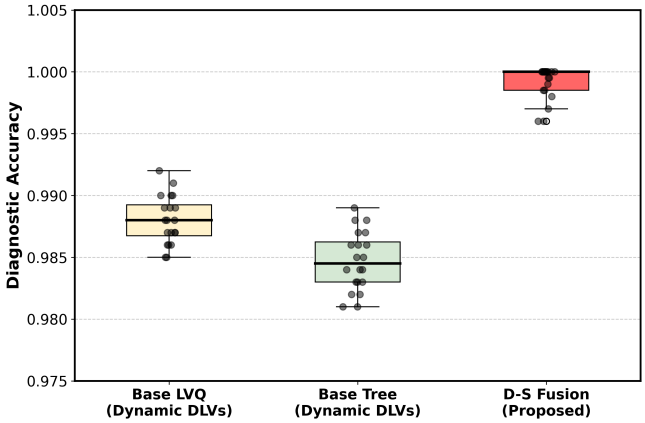


Figure 4. Internal ablation study of the PredVAR-driven classifiers and D-S fusion.

also show that the PCA-based fusion fluctuates more across repeated runs, whereas the PredVAR-based fusion remains close to the upper accuracy range. This observation further supports the use of dynamic features before the final decision-level fusion stage.

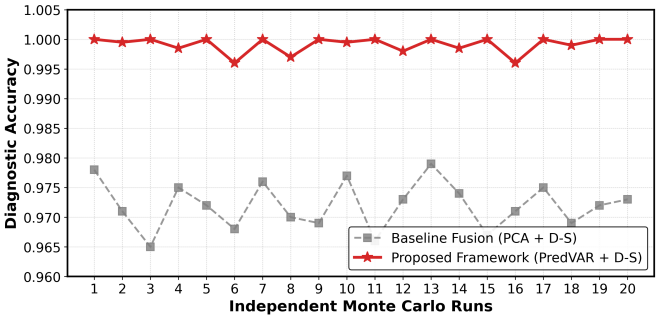


Figure 5. Diagnostic trajectories over 20 Monte Carlo runs.

3.3.2 Comparison with Recent Deep Learning Baselines

Recent neural baselines were also tested under the same mixed-load random-window protocol. In our implementation, the 1D-ResNet reached 99.07% average accuracy, the 1D-Transformer reached 99.99% average accuracy, and the ResNet-based transfer-learning baseline was also close to 100% under the same evaluation protocol. These results indicate that the CWRU random-window setting is a relatively saturated benchmark for high-capacity neural models, especially when sufficient labeled windows are available.

Table 1. Core comparison of the proposed framework against baseline methods.

Method Category	Model / Algorithm	Mean Accuracy	Std. Dev. (decimal)
Traditional ML	Static Features + SVM	98.61%	0.0028
Deep Learning	1D-CNN (End-to-End)	99.07%	0.0046
Baseline Fusion	PCA + LVQ	95.80%	0.0035
	PCA + DT	95.20%	0.0040
	PCA + D-S Fusion	**97.20%**	**0.0042**
Proposed Framework	PredVAR + LVQ	98.80%	0.0021
	PredVAR + DT	98.45%	0.0025
	PredVAR + D-S Fusion	**99.91%**	**0.0014**

For this reason, PredVAR + D-S is not presented as an accuracy winner over every recent neural classifier. Its contribution is different. The method provides a transparent dynamic feature path, uses only a compact DLV-based feature vector before LVQ/DT/D-S decision making, and allows the reliability of the generated evidence to be examined through the conflict coefficient K. The later analyses show that this design reduces evidence conflict, keeps online inference light, and performs well with limited labels.

3.4 Feature Separability and Classification Details

To visually assess the discriminability of the extracted features, t-SNE was used to map the testing features into a two-dimensional space. As shown in Figure 6, the PredVAR features form more compact clusters than the PCA features. In the PCA feature space, the three health states are generally separated, but the clusters are relatively loose and some boundaries remain close. In contrast, the PredVAR feature space shows clearer sub-clusters corresponding to different operating conditions. This indicates that the DLV-based features separate the three health states while retaining some operating-condition structure within each state. The t-SNE plot therefore illustrates class separability, but it does not by itself show that load-related variation has been removed.

The classification details are further shown by the confusion matrices in Figure 7. In the single test fold shown here, the PCA-based fusion produces 10 misclassifications, several of which occur between the Normal state and the Outer Race Fault class. This fold is better than the average PCA + D-S accuracy in Table 1 (97.20%), so it is used only to illustrate the type of errors rather than the average error level. The error type is consistent with the limitation of static statistical features: under mixed loads, a high-load normal signal may have an amplitude profile close to that of a low-load faulty signal.

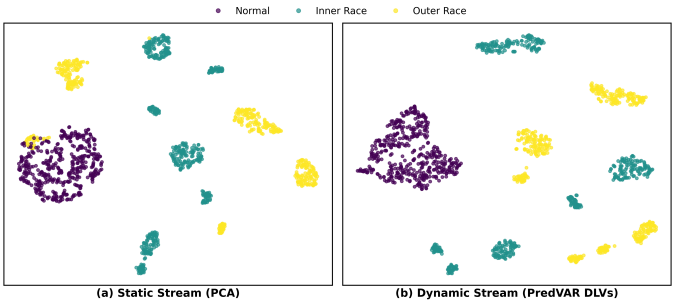


Figure 6. t-SNE visualization comparing feature separability between PCA-based static features and PredVAR-based DLV features on the representative test set.

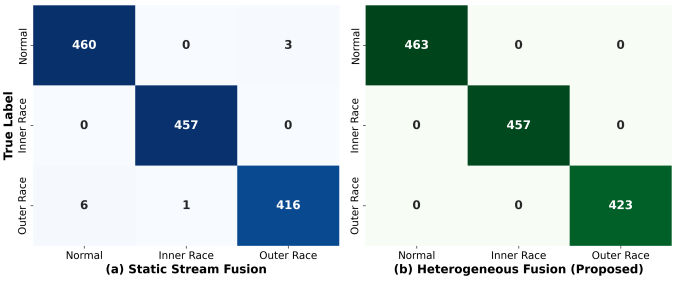


Figure 7. Confusion matrices of the PCA-based fusion baseline and the proposed PredVAR-based fusion framework in a single test fold.

The PredVAR-based fusion removes these errors in the shown fold and correctly classifies all 1343 testing samples. This result does not mean that every run is error-free, as the average performance is reported in Table 1. It shows, however, that the dynamic representation can reduce boundary ambiguity in the shown testing fold by focusing on predictable fault-related temporal patterns.

3.5 Mechanism Analysis via Prediction Error

To examine whether the extracted DLVs indeed contain predictable dynamics, the FPE of PredVAR was analyzed. When PredVAR is applied to

raw delay-embedded signals, the FPE converges to approximately 3×10^{-6}. This small value indicates that the latent variables have strong one-step predictability and contain regular dynamic components of the bearing signal.

As a negative control, PredVAR was also applied to static statistical feature vectors, which are treated as independent samples without temporal ordering. In this case, the FPE is close to 1.0, meaning that the model has little predictive power, as expected for data without temporal structure. This comparison supports applying PredVAR to delay-embedded raw sequences rather than to isolated statistical feature vectors. It shows that predictability requires temporal structure, but it does not by itself show that the extracted DLVs are specific to fault impacts.

3.6 Evidence Conflict Analysis

To directly evaluate the interaction between PredVAR and D-S fusion, the D-S conflict coefficient K in Eq. (17) was computed for the PCA + D-S and PredVAR + D-S frameworks. The same 20-run mixed-load protocol was used, and the conflict statistics are reported separately from the accuracy comparison in Table 2.

Table 2. Evidence conflict coefficient K under the main 20-run protocol.

Feature Source	Mean K	Std. of Run-Mean K	Reduction
PCA + D-S	0.0195	0.0038	–
PredVAR + D-S	0.0056	0.0057	71.3%

The mean conflict coefficient decreases from 0.0195 for PCA + D-S to 0.0056 for PredVAR + D-S, corresponding to a 71.3% reduction. This result gives direct evidence for the synergy described in Section 2.7: PredVAR improves the evidence source before D-S fusion, rather than only adding another preprocessing step. Note that the standard deviation of the run-mean K for PredVAR + D-S (0.0057) is not smaller than that for PCA + D-S (0.0038). The conclusion drawn from Table 2 is therefore limited to a lower average conflict level, not to a smaller run-to-run variation of K.

3.7 Hyperparameter Sensitivity Analysis

The default PredVAR parameters were further examined by a one-factor-at-a-time sensitivity study. The tested values were $m \in \{32, 64, 96\}$, $\tau \in \{1, 2, 4\}$, $s \in \{1, 2, 3\}$, and $\ell \in \{2, 4, 6\}$. When one parameter was varied, the other three were fixed at the default setting $m = 64$, $\tau = 1$, $s = 2$, and $\ell = 4$.

The accuracy range in Table 3 denotes the range of the 20-run mean accuracies obtained when the tested values of one parameter are varied.

The results indicate that the selected parameters are stable rather than finely tuned to a single setting. The default setting is close to the best observed performance while keeping the latent representation compact. In particular, using only two DLVs leads to clear underfitting, while $\ell = 4$ keeps most of the diagnostic information with a small feature dimension.

3.8 Computational Cost Analysis

The computational cost was considered in two stages: offline model construction and online inference. All computational-cost experiments were conducted on a Windows x64 workstation equipped with an Intel Core i7-12700H CPU, 24 GB RAM, and an NVIDIA GeForce RTX 3070 Ti Laptop GPU. The proposed PredVAR + D-S pipeline and the PCA + D-S baseline were implemented in Python 3.11.5 using NumPy 1.26.4, SciPy 1.17.1, and scikit-learn 1.8.0, and were executed on CPU only without GPU acceleration.

In our implementation, the offline setup time was 1.89 s for PCA + D-S and 24.33 s for PredVAR + D-S. The additional offline cost of PredVAR comes from identifying the dynamic projection matrix and the latent VAR model. This cost is paid during model construction and is not repeated during online diagnosis. The tested neural baselines also require an offline training stage before deployment.

The online stage is more relevant for practical diagnosis. After the model parameters have been fixed, PredVAR + D-S only requires delay embedding, matrix projection, statistical pooling of the DLVs, and lightweight LVQ/DT/D-S decision making. In the same local test environment, the online inference time of PredVAR + D-S was 0.818 ms per 2048-point window, compared with 0.610 ms for PCA + D-S. Since one 2048-point window at 12 kHz corresponds to 170.67 ms of signal acquisition time, both pipelines are much faster than real-time data acquisition.

From a scalability perspective, the online cost mainly depends on the window length, the embedding dimension m, and the number of retained DLVs ℓ. These operations follow a direct matrix-projection and feature-pooling process. Therefore, the main practical requirement is to complete the dynamic projection setup before deployment, after which the same projection can be reused for online monitoring.

Table 3. PredVAR hyperparameter sensitivity summary.

Parameter	Tested Values	Default Value	Accuracy Range	Observation
Embedding dimension m	32, 64, 96	64	99.82–99.94%	$m = 64$ gives near-best accuracy with moderate dimension.
Delay τ	1, 2, 4	1	98.91–99.94%	$\tau = 1$ gives near-best accuracy and preserves local temporal dynamics.
VAR order s	1, 2, 3	2	99.82–99.94%	$s = 2$ gives near-best accuracy and balances prediction ability and complexity.
Latent dimension ℓ	2, 4, 6	4	96.01–99.96%	$\ell = 2$ underfits; $\ell = 4$ is compact and near-best ($\ell = 6$ reaches the highest value, 99.96%).

3.9 Label-Efficiency Analysis

A further low-label experiment was conducted to check whether the proposed dynamic representation is useful when labeled fault data are limited. The number of labeled training windows per class was restricted, while the same test set was used for evaluation. Table 4 reports the result at 50 labeled windows per class.

Table 4. Label-efficiency comparison at 50 labeled windows per class.

Method	Accuracy
PredVAR + D-S	$97.17\% \pm 2.19\%$
PCA + D-S	$96.57\% \pm 0.82\%$
1D-ResNet	$96.09\% \pm 7.34\%$
1D-Transformer	$95.31\% \pm 2.22\%$
1D-CNN	$94.65\% \pm 3.08\%$

At this label budget, PredVAR + D-S achieves the highest mean accuracy among the compared methods. The margin over PCA + D-S (0.60 percentage points) is within the reported standard deviations, and PredVAR + D-S shows larger run-to-run variation than PCA + D-S at this budget (2.19% vs. 0.82%). This result therefore suggests a possible advantage of the proposed framework under limited-label conditions, rather than a conclusive one. A plausible reason is that the low-dimensional DLV representation reduces the amount of information that must be learned by the classifiers, so the model may remain effective when only a small number of labeled fault windows is available. This low-label result is also consistent with the full-data comparison: high-capacity neural models can be very strong when enough labeled data or pretrained representations are available, while PredVAR + D-S is more attractive when interpretability and label efficiency are important.

4 Conclusion

This paper proposed a PredVAR-driven evidence fusion framework for rolling bearing fault diagnosis. Instead of using static PCA features, the method extracts predictable DLVs from delay-embedded vibration signals and uses them as the evidence source for LVQ, DT, and D-S evidence fusion. The main findings are summarized as follows:

1. **Improved accuracy and stability:** Under the main 20-run mixed-load CWRU protocol, PredVAR + D-S achieves an average accuracy of 99.91% with a standard deviation of 0.0014. Compared with PCA + D-S, the dynamic representation improves both mean accuracy and run-to-run stability.
2. **Dynamic and more consistent evidence:** The FPE analysis shows that PredVAR extracts highly predictable DLVs from raw delay-embedded signals, whereas static statistical features contain little predictive structure. The t-SNE visualization, confusion matrices, and conflict analysis further show that these DLV-based features improve class separation and reduce the mean D-S conflict coefficient from 0.0195 to 0.0056.
3. **Practicality under limited labels:** In the experiment with 50 labeled windows per class, PredVAR + D-S obtains the highest mean accuracy among the compared methods. Its online inference time is also much shorter than the time required to acquire one vibration window, making the method suitable for continuous monitoring after the offline PredVAR model has been constructed.
4. **Position relative to neural baselines:** The mixed-load random-window CWRU protocol is close to saturated for modern neural classifiers. In the full-data setting, 1D-Transformer and ResNet-based transfer learning can achieve higher mean accuracy than PredVAR + D-S. Therefore, the proposed method is better viewed as a compact and interpretable dynamic-fusion approach, especially when evidence consistency, simple online inference, and limited labeled data are important.

The present study has limitations. The main protocol uses random splits of non-overlapping windows, so windows from the same recording can appear in both training and test sets. Only three health states are considered, and ball faults are not included. The reported results should therefore be read as evidence under this protocol, not as a guarantee of cross-condition generalization.

Future work will test the framework on datasets with stronger distribution shifts and explore how PredVAR-based dynamic features can be combined with neural models while retaining interpretability.

Data Availability Statement

The CWRU bearing data used in this study are publicly available from the Case Western Reserve University Bearing Data Center (https://engineering.case.edu/bearingdatacenter). Other data supporting the findings of this study are available from the corresponding author upon request.

Funding

This work was supported in part by the Hong Kong Research Grants Council under the General Research Fund (16206324,13300525) and in part by the Lingnan University under Grant SUG-010/2425 and Grant DR26F4. This work was also supported by the National Natural Science Foundation of China (NSFC) under Grant 62603322.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that ChatGPT-5 was used for language editing in the Methods section. The authors have carefully reviewed, revised, and verified the AI-assisted output and take full responsibility for the content of the manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Peng, B., Bi, Y., Xue, B., Zhang, M., & Wan, S. (2022). A survey on fault diagnosis of rolling bearings. *Algorithms, 15*(10), 347. [CrossRef]

[2] Zhao, J., Wang, W., Huang, J., & Ma, X. (2025). A comprehensive review of deep learning-based fault diagnosis approaches for rolling bearings: Advancements and challenges. *AIP Advances, 15*(2). [CrossRef]

[3] Lei, L., Li, W., Zhang, S., Wu, C., & Yu, H. (2025). Research progress on data-driven industrial fault diagnosis methods. *Sensors, 25*(9), 2952. [CrossRef]

[4] Jaber, A. A. (2024). Diagnosis of bearing faults using temporal vibration signals: a comparative study of machine learning models with feature selection techniques. *Journal of Failure Analysis and Prevention, 24*(2), 752-768. [CrossRef]

[5] Hua, L., Wu, X., Liu, T., & Li, S. (2022). The methodology of modified frequency band envelope kurtosis for bearing fault diagnosis. *IEEE Transactions on Industrial Informatics, 19*(3), 2856-2865. [CrossRef]

[6] Prieto-Moreno, A., Llanes-Santiago, O., & García-Moreno, E. (2015). Principal components selection for dimensionality reduction using discriminant information applied to fault diagnosis. *Journal of Process Control, 33*, 14-24. [CrossRef]

[7] Brito, L. C., Susto, G. A., Brito, J. N., & Duarte, M. A. V. (2021). Fault detection of bearing: An unsupervised machine learning approach exploiting feature extraction and dimensionality reduction. *Informatics, 8*(4), 85. [CrossRef]

[8] Wen, X., & Xu, Z. (2021). Wind turbine fault diagnosis based on ReliefF-PCA and DNN. *Expert Systems with Applications, 178*, 115016. [CrossRef]

[9] Yao, X., Li, S., & Hu, J. (2017). Improving rolling bearing fault diagnosis by DS evidence theory based fusion model. *Journal of Sensors, 2017*(1), 6737295. [CrossRef]

[10] Kohonen, T. (1995). Learning vector quantization. In *Self-Organizing Maps* (pp. 245-261). Springer. [CrossRef]

[11] Nova, D., & Estévez, P. A. (2014). A review of learning vector quantization classifiers. *Neural Computing and Applications, 25*(3), 511-524. [CrossRef]

[12] Quinlan, J. R. (2014). *C4.5: Programs for Machine Learning*. Elsevier. https://dl.acm.org/doi/abs/10.5555/583200

[13] Suthaharan, S. (2016). Decision tree learning. In *Machine Learning Models and Algorithms for Big Data Classification* (pp. 237-269). Springer. [CrossRef]

[14] Zhen, D., Guo, J., Xu, Y., Zhang, H., & Gu, F. (2019). A novel fault detection method for rolling bearings based on non-stationary vibration signature analysis. *Sensors, 19*(18), 3994. [CrossRef]

[15] Gong, T., Yang, J., Liu, S., & Liu, H. (2022). Non-stationary feature extraction by the stochastic response of coupled oscillators and its application in bearing fault diagnosis under variable speed condition. *Nonlinear Dynamics, 108*(4), 3839-3857. [CrossRef]

[16] Kasim, N. A., Mohamed, M. G., & Nuawi, M. Z. (2022). Non-Stationary vibratory signatures bearing fault detection using alternative novel kurtosis-based statistical analysis. *Journal of Applied Science and Process Engineering, 9*(1), 1139-1148. [CrossRef]

[17] Mo, Y., & Qin, S. J. (2025). Probabilistic reduced-dimensional vector autoregressive modeling with oblique projections. *Automatica, 180*, 112476.

[CrossRef]

[18] Dong, Y., & Qin, S. J. (2018). A novel dynamic PCA algorithm for dynamic data modeling and process monitoring. *Journal of Process Control, 67*, 1-11. [CrossRef]

[19] Yuan, R., Lv, Y., Lu, Z., Li, S., & Li, H. (2023). Robust fault diagnosis of rolling bearing via phase space reconstruction of intrinsic mode functions and neural network under various operating conditions. *Structural Health Monitoring, 22*(2), 846-864. [CrossRef]

[20] Zhang, Z., Liu, B., Liu, Y., & Zhang, H. (2022). Fault feature-extraction method of aviation bearing based on maximum correlation Rényi Entropy and Phase-Space Reconstruction Technology. *Entropy, 24*(10), 1459. [CrossRef]

[21] Xiao, F., Wen, J., Pedrycz, W., & Aritsugi, M. (2024). Complex evidence theory for multisource data fusion. *Chinese Journal of Information Fusion, 1*(2), 134-159. [CrossRef]

[22] Murphy, C. K. (2000). Combining belief functions when evidence conflicts. *Decision support systems, 29*(1), 1-9. [CrossRef]

[23] Li, W., Han, D., Dezert, J., & Yang, Y. (2024). Basic belief assignment determination based on radial basis function network. *Chinese Journal of Information Fusion, 1*(3), 175-182. [CrossRef]

[24] Wen, G., Lei, Z., Chen, X., & Huang, X. (2024). DS evidence theory and its application for fault diagnosis of machinery. In *New Generation Artificial Intelligence-Driven Diagnosis and Maintenance Techniques: Advanced Machine Learning Models, Methods and Applications* (pp. 325-341). Singapore: Springer Nature Singapore. [CrossRef]

[25] Li, L., Zhang, X., Wang, P., Chen, C., Ma, T., & Gao, S. (2025). A bearing fault diagnosis method for multi-sensors using cloud model and Dempster–Shafer evidence fusion. *Applied Sciences, 15*(21), 11302. [CrossRef]

[26] Case Western Reserve University Bearing Data Center. (2024). *Bearing Data Center: Seeded Fault Test Data*. Case Western Reserve University. Accessed May 18, 2026. https://engineering.case.edu/bearingdatacenter

[27] Liang, P., Wang, W., Yuan, X., Liu, S., Zhang, L., & Cheng, Y. (2022). Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved ResNet under noisy labels and environment. *Engineering Applications of Artificial Intelligence, 115*, 105269. [CrossRef]

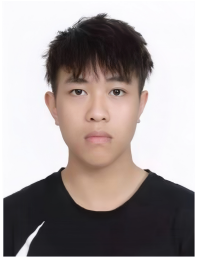
[28] Ni, Q., Ji, J. C., Halkon, B., Feng, K., & Nandi, A. K. (2023). Physics-Informed Residual Network (PIResNet) for rolling element bearing fault diagnostics. *Mechanical Systems and Signal Processing, 200*, 110544. [CrossRef]

[29] Hou, S., Lian, A., & Chu, Y. (2023). Bearing fault diagnosis method using the joint feature extraction of Transformer and ResNet. *Measurement Science and Technology, 34*(7), 075108. [CrossRef]

[30] Zhong, H., Lv, Y., Yuan, R., & Yang, D. (2022). Bearing fault diagnosis using transfer learning and self-attention ensemble lightweight convolutional neural network. *Neurocomputing, 501*, 765-777. [CrossRef]



Zhongxi Zhu received the B.Sc. degree in Computing Mathematics from the City University of Hong Kong and the M.Sc. degree in Mathematics from The Chinese University of Hong Kong. He is currently pursuing an M.Phil. degree in Industrial Data Science at Lingnan University, Hong Kong. His current research interests include data analytics and machine learning.



Weijun Lv received the Ph.D. degree in Control Science and Engineering from Guangdong University of Technology, Guangzhou, China. He was a visiting Ph.D. student at the Hong Kong University of Science and Technology. His research focuses on networked systems and state estimation.



Yanfang Mo received the B.E. degree in Automation from Zhejiang University, China, and the Ph.D. degree in Electronic and Computer Engineering from The Hong Kong University of Science and Technology. She has held several research positions, including Visiting Researcher at the University of California, Berkeley, Postdoc at City University of Hong Kong, and Research Assistant Professor at HKUST. She is currently an Assistant Professor in the Division of Industrial Data Science, School of Data Science at Lingnan University, Hong Kong. Her research interests lie at the intersection of control, optimization, data science, and operations research, with a focus on smart city development. She is the recipient of the best poster paper award at IEEE INFOCOM 2021.