Advanced Engineering & Precision Manufacturing
WISDOM ACADEMIC https://doi.org/10.63808/aepm.v2i3.424 Volume 2 , Issue 3

Research Article

A Method for Domain Generalization in Rolling Bearing Fault Diagnosis Based on Mamba and Causal Prototype Decoupling Constraints

Jiabing Zhou[1,*], Xiang Gu[1], Bo Zhang[1], Xinrui Yu[1], Yining Xu[1]
[1]School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing 102488, China
[*]**Correspondence:** Jiabing Zhou, 13846403521@163.com

Abstract

To address the insufficient generalization capability of rolling bearing fault diagnosis models under complex operating conditions such as variable rotational speed, variable load, and variable radial force, this paper proposes a domain generalization fault diagnosis method based on Mamba feature extraction and causal generalization loss. First, the raw vibration signals are standardized and segmented using a sliding window strategy. Then, the selective state space model Mamba is employed to model long-range temporal dependencies and local dynamic variations in fault impact signals. Subsequently, from the perspective of causal invariance, generalization constraints are constructed by treating stable representations related to fault categories as causal features, while regarding amplitude variations, noise disturbances, and speed fluctuations induced by operating-condition changes as non-causal factors. A causal generalization loss consisting of class-conditional causal prototype consistency loss and feature decoupling regularization is designed to enhance the model's adaptability to unseen operating conditions. Experiments are conducted on the Paderborn University (PU) and JNU datasets, where four operating conditions are regarded as four domains to construct cross-condition diagnosis tasks under a leave-one-condition-out evaluation protocol. The proposed method is compared with support vector machine (SVM), one-dimensional convolutional neural network (1D-CNN), one-dimensional residual network (ResNet1D), temporal convolutional network (TCN), correlation alignment (CORAL), domain-adversarial neural network (DANN), maximum mean discrepancy (MMD), and vanilla Mamba. Experimental results show that the proposed method achieves an average accuracy of 93.56% on the PU dataset and 96.28% on the JNU dataset, outperforming all comparison methods and further demonstrating its effectiveness.

Keywords

Rolling bearings, Fault diagnosis, Domain generalization, Causal prototypes, Feature decoupling

1. Introduction

Rolling bearings are critical support components in rotating machinery such as electric motors, fans, pumps, and CNC machine tools; their operating condition directly affects the reliability and safety of the equipment system. When damage occurs to the inner ring, outer ring, or rolling elements of a bearing, vibration signals typically exhibit periodic impacts, amplitude modulation, and non-stationary spectral changes. Consequently, intelligent fault diagnosis based on vibration signals has become a major research focus in the field of mechanical condition monitoring (Lessmeier et al., 2016; Paderborn University, Chair of Design and Drive Technology, n.d.; Randall & Antoni, 2011).

In recent years, deep learning methods have significantly advanced the automation of rolling bearing fault diagnosis. Convolutional neural networks can extract local impact patterns from raw one-dimensional signals, residual networks can alleviate the challenges of training deep models, and time-series models such as TCNs can enhance the ability to model sequence dependencies (Bai et al., 2018; He et al., 2016; Janssens et al., 2016; Lei et al., 2020; Zhang et al., 2018; Zhang et al., 2020). However, actual industrial equipment often operates under variable conditions over extended periods, resulting in significant distribution shifts between training and test data. For example, variations in rotational speed, load torque, radial load, and temperature can all alter the amplitude distribution, impact period, and spectral structure of vibration signals. Traditional supervised learning models are prone to learning spurious correlations specific to particular operating conditions; when these models are directly applied to unseen operating conditions, diagnostic accuracy may decline significantly.

For cross-condition diagnostic problems, transfer learning, domain adaptation, and domain generalization methods have been extensively studied (Ganin et al., 2016; Pan & Yang, 2010; Sun & Saenko, 2016; Wang et al., 2023; Zhao et al., 2024). Among these, domain adaptation typically requires unlabeled samples from the target domain during the training phase, whereas domain generalization trains models using only data from multiple source domains and requires the model to generalize directly to an unknown target domain, which better aligns with the practical application needs of equipment deployment where target operating conditions are unpredictable. Common methods include CORAL, which is based on statistical distribution alignment; DANN, which is based on domain adversarial learning; and MMD, which is based on maximum mean deviation. While these methods can mitigate operating condition differences between source domains to some extent, their feature extraction backbones largely rely on CNNs or conventional temporal networks, resulting in limited capabilities for modeling cross-period fault impacts in long vibration sequences. Recent studies on large-model-based machine monitoring, digital-twin-driven predictive maintenance, and machine-learning-based engineering optimization have also shown that robust fault diagnosis should consider not only diagnostic performance but also predictive maintenance requirements and engineering deployment conditions (Chen et al., 2025; Hasan et al., 2025; Lu et al., 2025; Lu et al., 2026; Wang et al., 2020).

Mamba is a selective state-space model proposed in recent years. It dynamically retains or forgets sequence information through an input-dependent selection mechanism and is capable of processing long sequences with approximately linear computational complexity (Gu & Dao, 2023; Gu et al., 2022). For bearing vibration signals, fault impacts exhibit both local abrupt changes and long-range dependencies that recur across cycles. Compared to simple convolutional architectures, Mamba is better suited for extracting fault patterns over extended time intervals; compared to Transformers, it has lower computational complexity and is more suitable for processing one-dimensional long-sequence signals. More specifically, local bearing defects generate transient impact responses, and defect-induced impacts repeatedly occur over multiple shaft rotation cycles. The input-dependent selective scanning mechanism of Mamba can preserve cross-cycle information without introducing the quadratic attention overhead of Transformers. Compared with TCNs with fixed receptive fields or recurrent networks, Mamba can adaptively emphasize signal segments containing fault impacts while suppressing steady-state background vibration, making it more suitable for modeling long one-dimensional vibration sequences.

Furthermore, causal learning provides a new interpretive framework for domain-generalized fault diagnosis. Based on the concept of causal invariance, the stable damage features that truly determine fault categories should remain relatively consistent across different environments or operating conditions, whereas non-causal disturbance features caused by rotational speed, load, and noise may vary with changes in the domain (Lv et al., 2022; Mahajan et al., 2021; Pearl, 2009; Peters et al., 2016). Therefore, if a model can enhance the representational consistency of the same fault category across different operating conditions and suppress dependence on operating condition disturbance features, it is expected to achieve better generalization capabilities in unknown domains. It should be specifically noted that the term "causal features" in this study is used in the sense of invariant representation learning inspired by causal invariance. It refers to predictive features that are related to fault mechanisms and remain stable across domains, rather than indicating that the model learns a complete structural causal model.

Based on the above analysis, this paper proposes a method for domain generalization in rolling bearing fault diagnosis based on Mamba and causal prototype-decoupling constraints. The main contributions of this paper include: First, constructing a Mamba feature extraction network for one-dimensional vibration signals to capture local impacts and long-range temporal dependencies; second, designing a causal generalization module comprising a class-conditional causal prototype consistency loss and a feature decoupling regularization term to enhance feature stability across operating conditions; third, validating the effectiveness of the method on the PU and JNU dataset through comparative and ablation experiments.

2. Method

2.1. Problem Definition

Suppose the training data consists of K source domains, with each source domain corresponding to a specific operating condition in the dataset. The set of source domains can be represented as:

$$D_s = D_1, D_2, \ldots, D_K, D_k = (x_i^k, y_i^k)_{i=1}^{n_k} \tag{1}$$

Here, x_i^k represents a vibration signal sample from the k-th source domain, and y_i^k represents the corresponding fault class label. The goal of the domain generalization task is to enable the model to achieve high diagnostic accuracy on the target domain D_t— which was not seen during training—by training the model using data from multiple source domains only.

From a causal perspective, fault category Y is primarily determined by the stable causal features Z_c, while environmental variables E—such as rotational speed, load torque, radial load, and noise—primarily influence the non-causal disturbance features Z_s. An ideal diagnostic model should classify faults using Z_c as much as possible and minimize its reliance on Z_s. This objective can be expressed as:

$$P(Y|Z_c, E_i) \approx P(Y|Z_c, E_j), i \neq j \tag{2}$$

Equation (2) indicates that the classification relationship based on fault causal features should remain relatively stable under different operating conditions.

2.2. Overall Framework

As shown in Figure 1, the method proposed in this paper primarily consists of a signal preprocessing module, a Mamba feature extractor, a causal-perturbation feature mapping module, a fault classifier, and a causal generalization constraint module. First, the raw vibration signal is Z-score normalized, and a sliding window is used to segment the continuous signal into sample sequences of fixed length; Second, the Mamba network is used to extract the fundamental temporal representation from the vibration signal; subsequently, the causal mapping head and perturbation mapping head decompose the fundamental representation into causal and perturbation representations; finally, the fault classifier performs class prediction based on the causal representation and constrains the feature space using class-conditional causal prototype consistency loss and feature decoupling regularization, thereby enhancing the model's generalization ability under unknown operating conditions.

Let the Mamba feature extractor be $G_f(\cdot)$, the causal mapping head be $\mathrm{h}_c(\cdot)$, the perturbation mapping head be $\mathrm{h}_s(\cdot)$, and the fault classifier be $G_y(\cdot)$. For an input vibration signal x, the forward computation process of the model is as follows:

$$z = G_f(x) \tag{3}$$

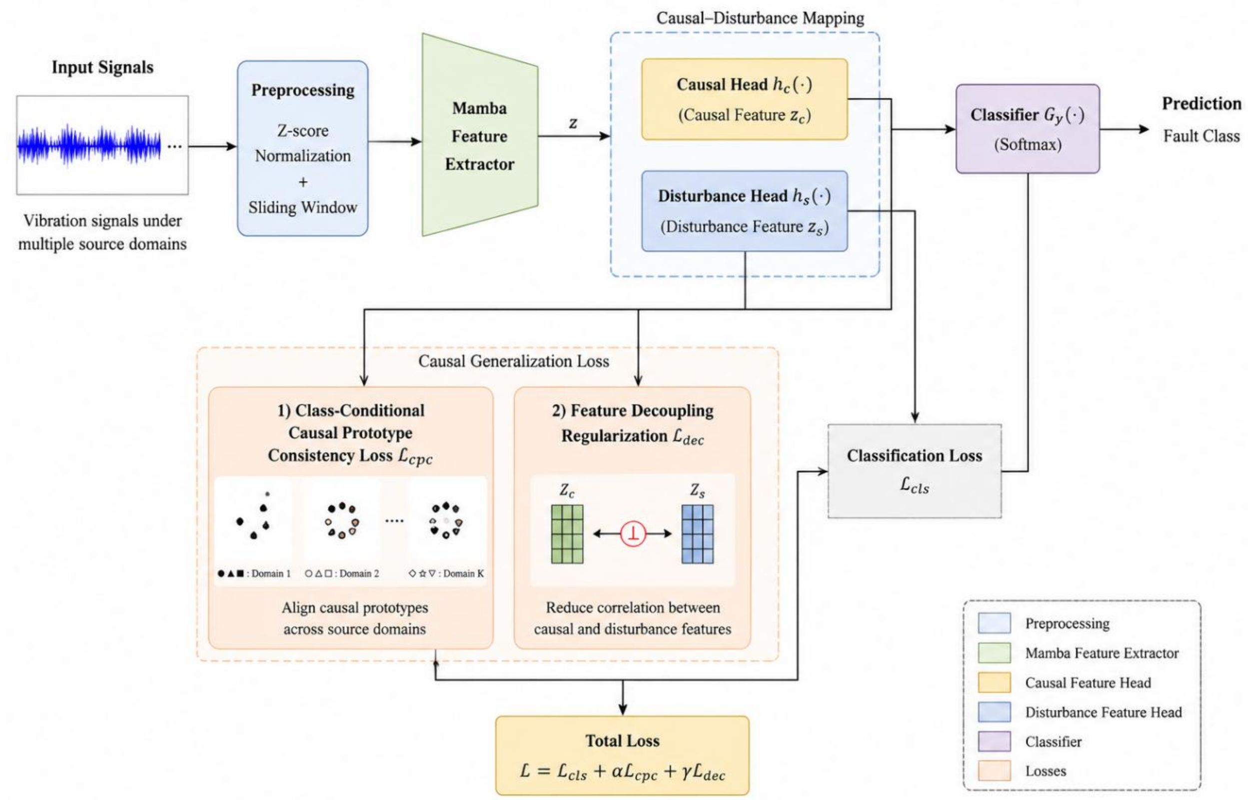
$$z_c = \mathrm{h}_c(z), \quad z_s = \mathrm{h}_s(z) \tag{4}$$

$$\hat{y} = G_y(z_c) \tag{5}$$

Here, z represents the basic time-series representation output by the Mamba feature extractor, z_c represents the causal representation, z_s represents the disturbance representation, and $\hat{y}$ represents the predicted probability of the fault class. Unlike directly using the mixed feature z for classification, this paper applies the classifier to z_c. The aim is to encourage the model to rely primarily on stable discriminative information related to the failure damage mechanism for diagnosis, thereby reducing the impact of operational disturbances such as rotational speed, load, and radial load on the classification results.

Figure 1

Framework



Note. G_f denotes the Mamba feature extractor; h_c and h_s denote the causal and disturbance mapping heads, respectively; G_y denotes the fault classifier; L_{cls} denotes classification loss; L_{cpc} denotes class-conditional causal prototype consistency loss; and L_{dec} denotes feature decoupling regularization.

It should be noted that the disturbance representation z_s is not directly used for fault classification but is incorporated into training via feature decoupling regularization to constrain the correlation between the causal representation and operational disturbance information. During the testing phase, the model relies on $G_f(\cdot)$, $h_c(\cdot)$, and $G_y(\cdot)$ to perform fault identification under unknown operating conditions.

2.3. Mamba Feature Extraction Module

At the core of Mamba lies a selective state-space mechanism. Unlike traditional state-space models with fixed state transition parameters, Mamba can dynamically generate state update parameters based on the input, thereby selectively modeling key information within the sequence. For one-dimensional vibration signals, this paper first employs a one-dimensional convolutional embedding layer to map local features:

$$H_0 = \mathrm{Conv1D}(x) \tag{6}$$

Here, H_0 represents the initial sequence representation of the input vibration signal after passing through a local convolution map. H_0 is then fed into a feature extraction network consisting of several layers of Mamba Blocks. The output of the lth layer of Mamba Blocks is expressed as:

$$H_l = H_{l-1} + \mathrm{MambaBlock}(\mathrm{LN}(H_{l-1})) \tag{7}$$

Here, LN(·)denotes the layer normalization operation, and the residual connection is used to mitigate the problem of gradient vanishing during the training of deep networks. Each Mamba Block primarily consists of a normalization layer, a selective state space layer, gated branches, and residual connections. This architecture is capable of preserving local impulse responses in vibration signals while modeling long-range dependencies between fault impulses across cycles, making it suitable for fault feature extraction from bearing vibration signals.

Finally, global average pooling is applied to the output sequence of the last Mamba Block, and a fully connected layer is used to obtain the basic fault representation:

$$z = \mathrm{FC}(\mathrm{GAP}(H_L)) \tag{8}$$

Here, GAP(·)denotes global average pooling, H_L denotes the output of the last Mamba Block, and z denotes the base temporal representation obtained.

2.4. Causal Generalization Module

To enhance the model's generalization ability to unknown operating conditions, this paper constructs a causal generalization module comprising the class-conditional causal prototype consistency loss L_{cpc} and the feature decoupling regularization L_{dec}. This module constrains the feature space in two ways: first, by aligning causal prototypes of the same fault category across different source domains, it promotes more consistent causal representations for faults of the same type under cross-condition scenarios; second, by reducing the correlation between causal representations and perturbation representations, it minimizes the interference of condition-specific information on fault classification results.

2.4.1. Class-Conditional Causal Prototype Consistency Loss

The same fault category typically exhibits identical or similar physical damage mechanisms under different operating conditions; therefore, its stable causal representation should have similar category centers. To this end, this paper constructs class-conditional causal prototypes across source domains and constrains the prototype distance for the same fault category across different source domains.

To avoid confusion between the category index and the subscript of the causal representation z_c, we use m to denote the failure category below. For the causal prototype of category m in the e-th source domain, it is defined as:

$$p_m^e = \frac{1}{N_m^e}\sum_{i:y_i^e=m} z_{c,i}^e = \frac{1}{N_m^e}\sum_{i:y_i^e=m} \mathrm{h}_c\left(G_f(x_i^e)\right) \tag{9}$$

Here, p_m^e denotes the causal prototype for category m in the e-th source domain, N_m^e denotes the number of samples for category m in the e-th source domain, and $z_{c,i}^e$ denotes the causal representation of the i-th sample in the e-th source domain.

After obtaining the class-conditional causal prototypes for each source domain, the class-conditional causal prototype consistency loss is defined as:

$$L_{cpc} = \frac{1}{C}\sum_{m=1}^{C} \frac{2}{K(K-1)}\sum_{1\le e_i<e_j\le K} \left\| p_m^{e_i} - p_m^{e_j} \right\|_2^2 \tag{10}$$

Here, C denotes the number of fault categories, and K denotes the number of source domains. This loss encourages causal prototypes of the same fault category to be closely aligned across different source operating conditions, thereby reducing intra-class distribution shifts caused by variations in rotational speed, load, and radial load.

It should be noted that, since samples from the unknown target domain are inaccessible during the domain generalization training phase, our method does not explicitly estimate the target domain prototype but instead aligns class-conditional causal prototypes across multiple source domains. This design is based on the following assumption: if the model can learn stable causal representations of the same fault type across multiple known source operating conditions, it is more likely to extract fault representations with similar discriminative significance in unknown target operating conditions.

Furthermore, the class-conditional causal prototype p_m^e serves solely as a regularization constraint during training to guide cross-domain alignment in the causal representation space; it is not directly used for prediction by the classifier. Fault classification is still performed by the classifier $G_y(\cdot)$ based on the causal representation z_c of a single sample. This approach prevents the model from relying on the target domain prototype during the testing phase, thereby adhering to the fundamental premise of domain generalization tasks where the target domain is unseen.

2.4.2. Feature Decoupling Regularization

Although the class-conditional causal prototype consistency loss helps ensure consistency in the causal representation of similar faults across different source domains, the base representation z may still contain disturbance information related to the operating conditions. To further reduce the confounding between fault causal features and operating condition disturbance features, this paper employs two mapping heads to decompose the base representation z into a causal representation z_c and a disturbance representation z_s:

$$z_c = \mathrm{h}_c(z), \quad z_s = \mathrm{h}_s(z) \tag{11}$$

In particular, $\mathrm{h}_c(\cdot)$ and $\mathrm{h}_s(\cdot)$ represent the causal feature mapping head and the perturbation feature mapping head, respectively. To reduce the correlation between the two, this paper introduces an orthogonal decoupling constraint:

$$L_{dec}=\left\|\frac{1}{B}\tilde{Z}_c^{\mathrm{T}}\tilde{Z}_s\right\|_F^2 \tag{12}$$

Here, B denotes the mini-batch size; $\tilde{Z}_c$ and $\tilde{Z}_s$ represent the causal representation matrix and the perturbation representation matrix, respectively, after in-batch centering and normalization; and $\|\cdot\|_F$ denotes the Frobenius norm. Equation (12) essentially constrains the correlation matrix between the causal representation and the perturbation representation to be close to zero, thereby reducing information redundancy between the two.

Through this constraint, the model can reduce the operational disturbance information contained in the causal representation, enabling the classifier to rely more on stable fault-discriminating features for identification. Compared to directly classifying the base representation z, this design explicitly mitigates the influence of operational factors at the feature level, thereby helping to improve the model's diagnostic generalization capability under unknown operating conditions.

2.5. Overall Optimization Goals

Cross-entropy loss is used for fault classification. Since this paper aims for the classifier to rely primarily on causal representations for fault identification, the classification prediction is based on z_c rather than the baseline mixed features z:

$$\hat{y}_i=G_y(z_{c,i}) \tag{13}$$

For a mini-batch of size B, the classification loss is defined as:

$$L_{cls}=-\frac{1}{B}\sum_{i=1}^{B}\sum_{m=1}^{C} y_{im}\log\left(\hat{y}_{im}\right) \tag{14}$$

Here, y_{im} is the one-hot label indicating that sample i belongs to class m, and $\hat{y}_{im}$ represents the model's predicted probability that sample i belongs to class m.

The final training objective is:

$$L=L_{cls}+\alpha L_{cpc}+\gamma L_{dec} \tag{15}$$

In particular, α and γ represent the weights of the class-conditional causal prototype consistency loss and the feature decoupling regularization, respectively. These three loss functions jointly optimize the model parameters: L_{cls} ensures that the causal representation possesses the ability to distinguish between fault categories; L_{cpc} promotes consistency in causal representations of the same fault type across different source operating conditions; and L_{dec} reduces the correlation between causal representations and operating condition perturbation representations.

This joint optimization has a clear mechanism of action. First, the classification loss L_{cls} prevents the degradation of causal representations, ensuring that z_c retains sufficient fault discrimination information; Second, the class-conditional causal prototype consistency loss L_{cpc} aligns representations of similar faults across different source domains at the category level, reducing intra-class distribution shifts across operating conditions; Finally, the feature decoupling regularization L_{dec} suppresses the dependence of causal representations on operating condition disturbance information. Through the synergistic interaction of these three components, the model is able to learn fault representations with greater stability across operating conditions, thereby enhancing diagnostic generalization capabilities under unknown operating conditions.

In the experiments of this paper, α=0.20 and γ=0.10 were set, and the impact of different weight values on model performance is further discussed in the parameter sensitivity analysis.

3. Experiments and Analysis

3.1. Introduction to the Dataset and Sample Construction

To comprehensively validate the generalization capability of the proposed method in bearing fault diagnosis across various operating conditions, this paper conducts experiments on the PU (Paderborn University) bearing dataset and the JNU bearing dataset. Both datasets contain vibration signals collected under a variety of operating conditions; however, they differ in terms of operating variables, data acquisition platforms, and fault distributions. Consequently, they allow for the evaluation of the model's adaptability to unknown operating conditions from different perspectives.

The PU dataset is provided by the KAt Bearing Data Center at the University of Paderborn in Germany and is primarily used for research on condition monitoring of rolling bearing damage in electric motor drive systems. The dataset simultaneously captures bearing vibration signals and motor current signals, while also recording auxiliary information such as rotational speed, load torque, radial force, and temperature. This paper uses only the vibration signals for fault diagnosis experiments. The PU dataset includes both healthy bearing conditions and real damage caused by artificial machining and accelerated life testing, thereby combining reproducible experimental conditions with representativeness of engineering failures. The four operating conditions are shown in Table 1.

Table 1

Operating Conditions for the PU Dataset

Domain ID	Operating Condition Name	RPM	Load torque (Nm)	Radial force (N)
A	N15_M07_F10	1500	0.7	1000
B	N09_M07_F10	900	0.7	1000
C	N15_M01_F10	1500	0.1	1000
D	N15_M07_F04	1500	0.7	400

Note. RPM = revolutions per minute; Nm = newton-meter; N = newton. Domains A–D represent four operating conditions in the PU dataset.

To ensure consistency in the feature space across different domains, this paper selects 10 bearing condition classes—denoted as C0–C9—that can be obtained under all four operating conditions as diagnostic features. For each class in each domain, 200 one-dimensional vibration samples are selected; thus, each domain contains 2,000 samples, and the four domains together contain 8,000 samples. The specific class settings are shown in Table 2.

Table 2

Diagnostic Category Settings for the PU Dataset

Category	Bearing Number	Fault Location/Type	Note
C0	K001	Health	Non-destructive reference bearing
C1	KA04	outer ring	Actual pitting damage
C2	KA15	outer ring	Plastic deformation/indentation
C3	KA16	outer ring	Actual pitting damage
C4	KA22	outer ring	Actual pitting damage
C5	KA30	outer ring	Plastic deformation/indentation
C6	KI04	inner ring	Actual pitting damage
C7	KI14	inner ring	Actual pitting damage
C8	KI16	inner ring	Actual pitting damage
C9	KI18	inner ring	Actual pitting damage

Note. C0–C9 denote the 10 diagnostic categories. KA and KI denote outer-ring and inner-ring bearing fault samples, respectively.

The JNU dataset contains bearing vibration signals under three speed conditions—600, 800, and 1000 r/min—as shown in Table 3 (Wang, n.d.), and can be used to evaluate the model's generalization ability across different speed ranges. Compared to the PU dataset, the operational variations in the JNU dataset are primarily reflected in changes in rotational speed; therefore, it serves as a complementary validation of the PU dataset's multi-condition variable scenarios. This paper treats the three rotational speeds as three independent domains and employs a leave-one-speed-out testing strategy for domain generalization experiments. In terms of diagnostic categories, this paper selects four types of bearing conditions—healthy state, inner ring failure, outer ring failure, and rolling element failure—as classification targets. For each speed domain, 200 one-dimensional vibration samples are selected for each category, resulting in a total of 800 samples per domain and 2,400 samples across all three domains. This setup ensures consistency in the number of categories and samples across different speed domains, thereby enhancing the comparability of cross-domain diagnostic results.

Table 3

Operating Conditions for the JNU Dataset

Domain ID	Operating Condition Name	RPM	Domain type
J1	600 r/min	600	Speed range
J2	800 r/min	800	Speed range
J3	1000 r/min	1000	Speed range

Note. RPM = revolutions per minute. J1–J3 represent the three rotational-speed domains in the JNU dataset.

Both datasets employ a domain generalization setup with one-domain testing. For the PU dataset, a three-domain training and one-domain testing strategy was adopted to construct four task pairs: A, B, C → D; A, B, D → C; A, C, D → B; and B, C, D → A; For the JNU dataset, a two-domain training and one-domain testing strategy was adopted, resulting in three task pairs: 600,800→1000, 600,1000→800, and 800,1000→600. Samples from the target domain were used only for the final test and did not participate in model training or parameter updates.

During the data preprocessing stage, the raw vibration signals are first Z-score normalized, then segmented into one-dimensional samples using a sliding window strategy. To determine an appropriate window length, we conducted preliminary experiments with different segment lengths, including 1024, 2048, and 4096. The results show that a window length of 2048 achieves the best overall performance on PU datasets, yielding higher classification accuracy compared to 1024 and 4096. Specifically, the average accuracies on the PU dataset are 91.48%, 93.56%, and 92.67% for window sizes of 1024, 2048, and 4096, respectively. Therefore, 2048 is selected as the default setting in this study. To prevent data leakage, the training domain and target domain are strictly separated from the operating conditions domain.

All experiments are conducted on a workstation equipped with an NVIDIA RTX 3090 GPU and were trained using the Pytorch 2.11, CUDA 12.1, Adam optimizer, with an initial learning rate of 0.001, a batch size of 64, and 100 training epochs. The source domain samples were split into a training set and a validation set in an 8:2 ratio; the validation set was used solely for model selection and convergence monitoring. The evaluation metric used was classification accuracy, calculated as follows:

$$Acc=N_{correct}/N_{total}\times 100\% \tag{16}$$

3.2. Comparative experiment

To validate the effectiveness of the proposed method, this paper compares it with traditional machine learning models, typical deep fault diagnosis models, domain generalization models, and sequence modeling models, including Support Vector Machines (SVM), one-dimensional Convolutional Neural Networks (1D-CNN), one-dimensional Residual Networks (ResNet1D), Time-Convolutional Networks (TCN), Correlation Alignment (CORAL), Domain Adversarial Neural Networks (DANN), Maximum Marginal Disparity (MMD), and standard Mamba. The SVM uses common time-domain statistical features for classification; 1D-CNN, ResNet1D, and TCN serve as typical deep time-series diagnostic models; CORAL, DANN, and MMD enhance feature consistency through source-domain distribution constraints; Standard Mamba employs only the Mamba feature extractor and cross-entropy loss, without introducing causal prototype consistency loss or feature decoupling regularization. All deep models use the same number of training epochs, batch size, and learning rate settings to ensure a fair comparison.

Table 4

Accuracy of Different Methods on the PU Cross-Condition Task (%)

Method	A,B,C→D	A,B,D→C	A,C,D→B	B,C,D→A	average
SVM	76.35	78.42	70.86	81.15	76.70
1D-CNN	84.62	86.38	80.54	88.20	84.94
ResNet1D	87.15	88.72	83.96	90.31	87.54
TCN	86.73	89.10	84.42	90.05	87.58
CORAL	88.96	90.24	86.11	91.43	89.19
DANN	89.58	91.02	86.74	92.10	89.86
MMD	88.27	89.61	85.32	90.78	88.50
Mamba	91.84	93.02	89.35	94.13	92.09
Ours	93.26	94.54	91.62	94.83	93.56

Note. SVM = support vector machine; 1D-CNN = one-dimensional convolutional neural network; ResNet1D = one-dimensional residual network; TCN = temporal convolutional network; CORAL = correlation alignment; DANN = domain-adversarial neural network; MMD = maximum mean discrepancy. The domains before and after the arrow indicate the training and testing domains, respectively. Ours denotes the proposed method.

Table 5

Accuracy of Different Methods on the JNU Cross-Speed Task (%)

Method	600,800→1000	600,1000→800	800,1000→600	average
SVM	84.26	85.14	84.02	84.47
1D-CNN	91.36	92.08	90.87	91.44
ResNet1D	94.12	94.85	94.14	94.37
TCN	94.68	95.10	94.68	94.82
CORAL	94.83	95.36	94.77	94.99
DANN	95.02	95.54	95.08	95.21
MMD	94.56	95.21	94.63	94.80
Mamba	95.26	95.74	95.23	95.41
Ours	96.05	96.61	96.18	96.28

Note. The speeds before and after the arrow indicate the training and testing domains, respectively.

As shown in Table 4, the average accuracy of the traditional SVM method on the PU dataset is only 76.70%, which is significantly lower than that of deep learning methods, indicating that manually extracted statistical features struggle to fully capture failure modes under complex operating conditions. The average accuracy rates of 1D-CNN, ResNet1D, and TCN were 84.94%, 87.54%, and 87.58%, respectively, indicating that deep models can automatically extract more discriminative time-series fault features. CORAL, DANN, and MMD outperformed standard deep models, suggesting that distribution constraints across source domains can mitigate feature shifts caused by operating condition variations to some extent. The average accuracy of standard Mamba reached 92.09%, surpassing that of CNN, ResNet1D, and TCN, indicating that selective state-space models are more effective at modeling long-range dependencies and impulse responses in vibration signals. The method proposed in this paper achieved the highest accuracy across all four sets of PU cross-operating condition tasks, with an average accuracy of 93.56%, representing a 1.47 percentage point improvement over standard Mamba.

Looking at the specific tasks in the PU, the overall accuracy for A, C, and D → B is lower than for other tasks. This is primarily because the B domain corresponds to a speed of 900 r/min, while most of the other source domains are at 1500 r/min. The difference in speed alters the fault impulse intervals and spectral energy distribution, thereby increasing the difficulty of cross-domain diagnosis. The proposed method still achieves an accuracy of 91.62% on this challenging task, demonstrating that the consistency of conditional causal prototypes and feature decoupling constraints can enhance the model's adaptability to unknown speed conditions.

As shown in Table 5, the overall accuracy of all methods on the JNU dataset is higher than that on the PU dataset, indicating that the inter-domain variations in the JNU cross-speed task are relatively concentrated, making diagnosis slightly less challenging. Nevertheless, a clear hierarchy of performance still exists among the different methods. The average accuracy of SVM is 84.47%, which is improved to 91.44% by 1D-CNN, while ResNet1D, TCN, CORAL, DANN, and MMD all achieve further improvements to the 94%–95% range. The average accuracy of the standard Mamba model is 95.41%, and our method further improves this to 96.28%, achieving the best results across all three cross-speed tasks. These results indicate that the proposed causal prototype–decoupling constraint is not only applicable to PU multi-operating-condition scenarios but also remains stable and effective in JNU cross-speed scenarios.

By combining the results from the two datasets, it is evident that the advantages of the method proposed in this paper stem not only from the Mamba feature extractor itself, but also from the synergistic effect of long-sequence modeling capabilities and causal generalization constraints. Mamba enhances the expressive power of the underlying temporal representations, while the conditional causal prototype consistency loss and feature decoupling regularization further guide the model to learn cross-domain stable fault discrimination information, thereby improving diagnostic performance under unknown operating conditions.

To further evaluate the computational efficiency of the proposed method, we compare the number of parameters and inference latency with representative baseline models, including 1D-CNN, ResNet1D, TCN, and standard Mamba. All experiments are conducted under the same hardware environment.

Table 6

Computational Complexity Comparison of Different Methods

Method	Params (M)	FLOPs (G)	Inference Time (ms/sample)
1D-CNN	2.41	0.86	2.31
ResNet1D	4.73	1.52	3.12
TCN	3.28	1.11	2.85
Mamba	2.96	0.74	1.92
Ours	3.18	0.79	2.05

Note. Params = number of model parameters; M = million; FLOPs = floating-point operations; G = billion; ms/sample = milliseconds per sample. Ours denotes the proposed method.

The results in Table 6 show that although the proposed method introduces additional causal generalization modules, the overall computational cost remains comparable to standard Mamba and is significantly lower than CNN-based deep architectures. This demonstrates that the proposed method maintains strong practical feasibility for real-time fault diagnosis applications.

3.3. Ablation Experiment

To further validate the effectiveness of each component module, this paper conducts ablation experiments on a set of representative cross-operating-condition scenarios in the PU dataset. Specifically, the A, C, D→B as the ablation setup. In this task, the target domain B operates at 900 r/min, while most operating conditions in the source domain operate at 1500 r/min. This task exhibits a significant difference in the speed range, making it particularly suitable for evaluating the model's generalization ability under challenging cross-condition conditions. The ablation experiments were not trained or parameter-tuned on the target domain; all models employed the same training strategy as in the comparison experiments.

Table 7

Results of Ablation Experiments on the PU Dataset Using Domains A, C, and D for Training and Domain B for Testing

Model Settings	Mamba Feature Extraction	L_{cpc}	L_{dec}	Accuracy (%)
1D-CNN Baseline	×	×	×	80.54
Mamba	√	×	×	89.35
Mamba + L_{cpc}	√	√	×	90.43
Mamba + L_{dec}	√	×	√	90.18
Ours	√	√	√	91.62

Note. Lcpc = class-conditional causal prototype consistency loss; Ldec = feature decoupling regularization. A check mark indicates that the component was included, whereas a multiplication sign indicates that it was excluded. Ours denotes the proposed complete model.

As shown in Table 7, replacing the 1D-CNN baseline with the Mamba feature extractor increased the accuracy from 80.54% to 89.35%, indicating that Mamba's ability to model long-range dependencies and periodic impact patterns in vibration signals significantly improves fault recognition performance. Adding a class-conditional causal prototype consistency loss to the standard Mamba model further improved the accuracy to 90.43%, indicating that constraining the consistency of prototype centers for the same fault class across different source domains helps reduce distribution shifts within the same fault class across operating conditions.

After adding the feature decoupling regularization alone, the accuracy reached 90.18%, indicating that reducing the correlation between causal representations and operating condition disturbance representations can decrease the model's reliance on information from unstable operating conditions. When L_{cpc} and L_{dec} are used simultaneously, the model achieves a maximum accuracy of 91.62%. This indicates that cross-domain homogeneity aggregation and causal/perturbation decoupling have complementary effects: the former enhances the cross-domain consistency of similar faults at the category level, while the latter suppresses the interference of operating condition perturbation information on classification boundaries at the feature level.

In addition, we further monitored the training process to examine whether the joint optimization of multiple losses would slow down convergence or lead to unstable training. Under the same learning rate, batch size, and number of training epochs, all model variants converged normally, and no obvious oscillation or non-convergence was observed. This is because the cross-entropy loss remains the dominant optimization objective for maintaining class discriminability, while the class-conditional causal prototype consistency loss and feature decoupling regularization act as auxiliary constraints with moderate weights. Therefore, the joint optimization of the three losses does not introduce noticeable convergence instability; instead, it improves cross-condition generalization when the loss weights are properly balanced.

3.4. Parameter Sensitivity Analysis

To analyze the impact of the weights in the causal generalization module on model performance, this paper adjusts the weight α of the class-conditional causal prototype consistency loss and the weight γ of the feature decoupling regularization, and records the average accuracy on the PU and JNU datasets, respectively. The experimental results are shown in Table 8.

Table 8

Effect of Causal Generalization Module Weights on Accuracy

α	γ	PU Average Accuracy / %	JNU Average Accuracy / %
0	0	92.09	95.41
0.05	0.02	92.74	95.73
0.10	0.05	93.21	96.02
0.20	0.10	93.56	96.28
0.40	0.20	92.88	95.94

Note. α and γ denote the weights assigned to the class-conditional causal prototype consistency loss and feature decoupling regularization, respectively. PU = Paderborn University; JNU = Jiangnan University.

As shown in Table 8, when $\alpha = 0$ and $\gamma = 0$, the model degenerates into the standard Mamba, achieving average accuracy rates of 92.09% and 95.41% on the PU and JNU datasets, respectively. As α and γ are gradually increased, the model's accuracy on both datasets first increases and then decreases, achieving optimal performance when $\alpha = 0.20$ and $\gamma = 0.10$. These results demonstrate that appropriate

https://doi.org/10.63808/aepm.v2i3.424 Volume 2 , Issue 3

causal prototype consistency constraints and feature decoupling constraints can effectively enhance the model's cross-domain stability.

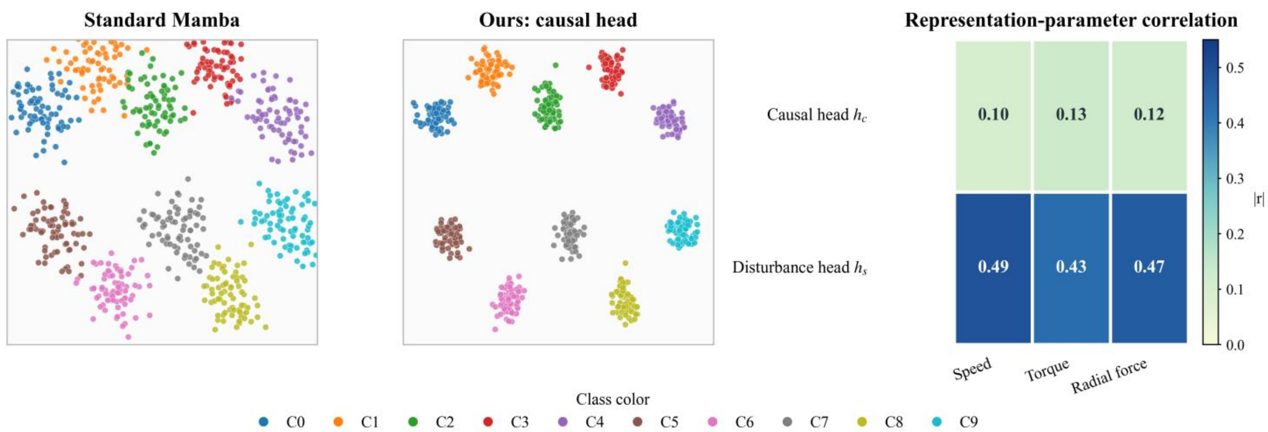
As the weights were further increased to α=0.40 and γ=0.20, the accuracy on both datasets declined. This suggests that overly strong cross-domain consistency and decoupling constraints may compress class discriminative information, causing the decision boundaries between different fault categories to become excessively smooth. Therefore, the causal generalization module needs to strike a balance between domain invariance and class separability. This paper adopts α = 0.20 and γ = 0.10 as the final parameter settings, a combination that performs stably on both the PU and JNU datasets.

3.5. T-SNE Visualization Analysis

To visually analyze the impact of the proposed method on the structure of the feature space, this paper further employs T-SNE to visualize the learned features in two dimensions. The A, C, D→B task is used as a representative case for the feature visualization analysis. Specifically, we extract the base features from the standard Mamba model and the causal head features from the proposed method, and map samples from different fault categories and source domains onto a two-dimensional space for comparison.

Figure 2

Feature visualization and decoupling analysis: (a) Standard Mamba features; (b) Causal head features from the proposed method; (c) Absolute Pearson correlation coefficients between learned representations and operating parameters



Note. t-SNE = t-distributed stochastic neighbor embedding. C0–C9 denote the 10 fault categories. Panels (a) and (b) show the features learned by standard Mamba and the proposed causal head, respectively; panel (c) shows the absolute Pearson correlation coefficients between learned representations and operating parameters.

As shown in Figure 2, although the standard Mamba can form a certain degree of class clustering, samples from the same class remain significantly dispersed across different operating conditions, indicating that the base representation still retains a significant amount of operating-condition-related noise. In contrast, the causal head features of the method proposed in this paper exhibit tighter intra-class clustering and clearer inter-class boundaries, indicating that the class-conditional causal prototype consistency loss effectively promotes feature alignment across different source domains for the same fault category.

Furthermore, as evidenced by the absolute Pearson correlation coefficients between the characteristics and operating parameters, the disturbance component exhibits a significantly higher average correlation with rotational speed, load torque, and radial force than the causal component. This indicates that the disturbance component tends to retain variations related to operating conditions, whereas the causal component is able to reduce dependence on operating variables. From a visualization perspective, this result supports the effectiveness of the method proposed in this paper in decoupling stable fault information from operating condition disturbance information.

For the JNU dataset, cross-domain variations primarily stem from changes in rotational speed. The experiments observed trends consistent with those of the PU dataset: while a certain degree of cross-rotational-speed offset still exists in the feature space of the standard Mamba model, the introduction of the causal generalization module further enhances the cross-domain aggregation of samples with similar faults. A comprehensive analysis of the visualization results from both datasets demonstrates that the proposed method not only improves numerical accuracy but also enhances cross-domain consistency in the representation of similar faults at the feature space level.

4. Conclusion

This paper proposes a domain generalization method for rolling bearing fault diagnosis based on Mamba and causal prototype decoupling constraints, and validates it across operating conditions using two public bearing datasets: PU and JNU. The method utilizes Mamba to model long-range dependencies and local impact patterns in one-dimensional vibration signals, while designing a causal generalization module comprising a class-conditional causal prototype consistency loss and a feature decoupling regularization term, enabling the model to learn more stable causal representations of faults.

Experimental results show that the proposed method achieved an average accuracy of 93.56% on the PU one-out-of-four-set test task and 96.28% on the JNU three-set cross-speed task, outperforming comparison methods such as SVM, 1D-CNN, ResNet1D, TCN, CORAL, DANN, MMD, and standard Mamba. Ablation experiments further demonstrate that the Mamba feature extractor, class-conditional causal prototype consistency loss, and feature decoupling regularization all play a positive role in enhancing cross-condition generalization performance. Parameter sensitivity analysis indicates that moderate causal generalization constraints can achieve a good balance between domain invariance and class separability. T-SNE visualization results further validate that our method enhances the cross-domain clustering of similar fault features and reduces the dependence of causal representations on operating condition perturbations.

Future work will focus on the following areas: First, validating the applicability of the proposed method using more real-world industrial data and cross-device tasks; second, developing a Mamba fault diagnosis model under multi-source sensor signal fusion; third, further elucidating the correspondence between causal features and bearing failure mechanisms through interpretability analysis; and fourth, exploring online prototype adaptation strategies for unlabeled streaming data and model compression schemes for edge deployment.

Acknowledgments

The authors have no additional acknowledgments to declare.

Author contributions

Conceptualization, Jiabing Zhou and Xiang Gu; methodology, Jiabing Zhou, Xiang Gu, and Bo Zhang; software, Xiang Gu and Bo Zhang; validation, Bo Zhang, Xinrui Yu, and Yining Xu; formal analysis, Xiang Gu and Xinrui Yu; investigation, Xiang Gu, Bo Zhang, Xinrui Yu, and Yining Xu; resources, Jiabing Zhou; data curation, Bo Zhang, Xinrui Yu, and Yining Xu; writing—original draft preparation, Xiang Gu and Bo Zhang; writing—review and editing, Jiabing Zhou, Xiang Gu, Xinrui Yu, and Yining Xu; visualization, Xiang Gu and Xinrui Yu; supervision, Jiabing Zhou; project administration, Jiabing Zhou; funding acquisition, Jiabing Zhou. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the 2026 Undergraduate Innovation and Entrepreneurship Training Program of Beijing Technology and Business University, grant number S202610011002.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Conflict of interest

The authors declare that there is no conflict of interest.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

All authors have given their consent.

Additional information

Received: 22 May 2026

Accepted: 18 June 2026

Published online: 16 August 2026

Publisher's Note

Wisdom Academic Press Ltd. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Arjovsky, M., Bottou, L., Gulrajani, I., & Lopez-Paz, D. (2019). Invariant risk minimization [Preprint]. arXiv. https://arxiv.org/abs/1907.02893

[2] Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [Preprint]. arXiv. https://arxiv.org/abs/1803.01271

[3] Chen, X., Lei, Y., Li, Y.-F., Parkinson, S., Li, X., Liu, J., Lu, F., Wang, H., Wang, Z., Yang, B., Ye, S., & Zhao, Z. (2025). Large models for machine monitoring and fault diagnostics: Opportunities, challenges, and future direction. *Journal of Dynamics, Monitoring and Diagnostics, 4*(2), 76–90. https://doi.org/10.37965/jdmd.2025.832

[4] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of Machine Learning Research, 17*(59), 1–35. https://www.jmlr.org/papers/v17/15-239.html

[5] Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces [Preprint]. arXiv. https://arxiv.org/abs/2312.00752

[6] Gu, A., Goel, K., & Ré, C. (2022). Efficiently modeling long sequences with structured state spaces [Preprint]. arXiv. https://arxiv.org/abs/2111.00396

[7] Hasan, M., Deng, Z., Redonnet, S., & Sanusi, B. M. (2025). Aerodynamic optimization of box-wing planform through machine learning integration. *Transactions of Nanjing University of Aeronautics and Astronautics, 42*(6), 789–800. https://doi.org/10.16356/j.1005-1120.2025.06.006

[8] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90

[9] Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., Van de Walle, R., & Van Hoecke, S. (2016). Convolutional neural network based fault detection for rotating machinery. *Journal of Sound and Vibration, 377*, 331–345. https://doi.org/10.1016/j.jsv.2016.05.027

[10] Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing, 138*, Article 106587. https://doi.org/10.1016/j.ymssp.2019.106587

[11] Lessmeier, C., Kimotho, J. K., Zimmer, D., & Sextro, W. (2016). 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, 3*(1). https://doi.org/10.36001/phme.2016.v3i1.1577

[12] Lu, Q., Cheng, L., Zhu, D., & Li, M. (2026). Fault prediction method towards gearbox based on digital twin and deep transfer learning. *Engineering Research Express, 8*(8), Article 085506. https://doi.org/10.1088/2631-8695/ae5b41

[13] Lu, Q., Li, M., & Huang, X. (2025). Digital twin and data-driven remaining useful life prediction of gearbox. *IEEE Access, 13*, 111614–111627. https://doi.org/10.1109/ACCESS.2025.3583313

[14] Lv, F., Liang, J., Li, S., Zang, B., Liu, C. H., Wang, Z., & Liu, D. (2022). Causality inspired representation learning for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8046–8056). IEEE. https://doi.org/10.1109/CVPR52688.2022.00788

[15] Mahajan, D., Tople, S., & Sharma, A. (2021). Domain generalization using causal matching. In *Proceedings of the 38th International Conference on Machine Learning* (pp. 7313–7324). PMLR. https://proceedings.mlr.press/v139/mahajan21a.html

[16] Paderborn University, Chair of Design and Drive Technology. (n.d.). *Bearing DataCenter*. Retrieved July 29, 2026, from https://mb.uni-paderborn.de/kat/forschung/bearing-datacenter

[17] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

[18] Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.

[19] Peters, J., Bühlmann, P., & Meinshausen, N. (2016). Causal inference using invariant prediction: Identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 78*(5), 947–1012. https://doi.org/10.1111/rssb.12167

[20] Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—A tutorial. *Mechanical Systems and Signal Processing, 25*(2), 485–520. https://doi.org/10.1016/j.ymssp.2010.07.017

[21] Sun, B., & Saenko, K. (2016). Deep CORAL: Correlation alignment for deep domain adaptation. In G. Hua & H. Jégou (Eds.), *Computer vision—ECCV 2016 workshops* (pp. 443–450). Springer. https://doi.org/10.1007/978-3-319-49409-8_35

[22] Wang, B., Lei, Y., Li, N., & Li, N. (2020). A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability, 69*(1), 401–412. https://doi.org/10.1109/TR.2018.2882682

[23] Wang, C. (n.d.). JNU bearing dataset [Data set]. GitHub. Retrieved July 29, 2026, from https://github.com/ClarkGableWang/JNU-Bearing-Dataset

[24] Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., Wang, L., Chen, Y., Zeng, W., & Yu, P. S. (2023). Generalizing to unseen domains: A survey on domain generalization. *IEEE Transactions on Knowledge and Data Engineering, 35*(8), 8052–8072. https://doi.org/10.1109/TKDE.2022.3178128

[25] Zhang, S., Zhang, S., Wang, B., & Habetler, T. G. (2020). Deep learning algorithms for bearing fault diagnostics—A comprehensive review. *IEEE Access, 8*, 29857–29881. https://doi.org/10.1109/ACCESS.2020.2972859

[26] Zhang, W., Li, C., Peng, G., Chen, Y., & Zhang, Z. (2018). A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing, 100*, 439–453. https://doi.org/10.1016/j.ymssp.2017.06.022

[27] Zhao, C., Zio, E., & Shen, W. (2024). Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study. *Reliability Engineering & System Safety, 245*, Article 109964. https://doi.org/10.1016/j.ress.2024.109964