Journal of Dynamics, Monitoring and Diagnostics, 2026, 5, 49-63
https://doi.org/10.37965/jdmd.2026.1087

ISTP
RESEARCH ARTICLE

A Dual-Component Elastic Adaptation Network for Rotating Machinery Incremental Fault Diagnosis under Variable Operating Conditions

Yan Zhang,[1] Changqing Shen,[1] Xiaofen Ye,[2] Liang Chen,[3] Juanjuan Shi,[1] and Zhongkui Zhu[1]

[1]School of Rail Transportation, Soochow University, Suzhou, P.R. China
[2]CRRC Qishuyan Locomotive & Rolling Stock Technology Research Institute Co., Ltd., Changzhou, China
[3]School of Mechanical and Electric Engineering, Soochow University, Suzhou, P.R. China

(Received 16 December 2025; Revised 05 January 2026; Accepted 16 January 2026; Published online 21 January 2026)

Abstract: In practical industrial environments, the data distribution of rotating machinery drifts as operating conditions vary, causing a marked deterioration in the performance of traditional fault diagnosis methods that rely on the assumption of identical distributions. Incremental learning provides a promising pathway to address dynamic operating conditions. However, existing approaches typically depend on replaying historical data and still struggle to strike a balance between stability and plasticity. To overcome these limitations, this paper proposes a dual-component elastic adaptive network (DCEAN) designed for incremental fault diagnosis of rotating machinery under varying working conditions. The proposed framework operates without access to previous data and simultaneously achieves knowledge retention and feature correction. Specifically, a sensitive parameter constraint (SPC) mechanism is introduced to curb excessive updates to parameters identified as critical, thereby stabilizing previously learned knowledge. In parallel, a feature drift self-calibration (FDSC) mechanism is employed to estimate and compensate for distribution shifts induced by condition variations, promoting consistency of feature representations across domains. Through the coordinated action of these two mechanisms, DCEAN establishes an incremental learning paradigm that harmonizes stability with adaptability. Two case studies demonstrate that the proposed method delivers superior diagnostic performance in variable operating environments, underscoring its robustness and effectiveness.

Keywords: feature drift; incremental learning; intelligent fault diagnosis; rotating machinery; variable operating conditions

I. INTRODUCTION

Rotating machinery constitutes critical equipment for power transmission and energy conversion within industrial production processes, with its operational stability and reliability directly impacting the efficient and safe functioning of entire industrial systems [1,2]. In such mechanical devices, rolling bearings serve as the core components supporting the operation of rotating parts, playing a crucial role. Extensive studies have shown that damage or failure of these critical parts is one of the primary causes of machinery downtime [3–5]. Therefore, achieving high-precision fault diagnosis for bearings and other fundamental components is crucial not only for ensuring long-term stable operation of equipment but also for supporting safe and intelligent production in modern industrial environments [6].

In recent years, rapid advances in sensing technologies and artificial intelligence have driven substantial interest in data-driven intelligent fault diagnosis (IFD) methods within the mechanical fault diagnosis community [7,8]. Early IFD approaches, such as support vector machines, decision trees, and k-nearest neighbors [9,10], rely heavily on the quality of manually designed features. These shallow models typically require significant effort from domain experts for feature construction, selection, and transformation. However, they struggle to effectively model unstructured, high-dimensional, or complex data, ultimately limiting diagnostic performance. To overcome these challenges, researchers have increasingly shifted toward deep learning (DL) paradigms, which offer end-to-end feature learning capabilities. Sun *et al.* [11] developed a lightweight and efficient Transformer architecture that enhances diagnostic accuracy while substantially reducing computational cost. Lu *et al.* [12] proposed a hybrid DL framework combining convolutional neural networks (CNNs) and transformers to handle complex operating conditions in rotating machinery fault diagnosis. Yu *et al.* [13] designed a CNN-GAN-based diagnostic framework and demonstrated its effectiveness in few-sample fault diagnosis scenarios.

Collectively, these studies highlight a clear trend: DL-based approaches have promoted a shift from expert-crafted features to autonomous feature learning and intelligent decision-making, providing strong technical support for the safe, stable, and efficient operation of modern industrial equipment. However, most existing DL-based diagnostic methods assume that training and testing data are drawn from identical operating conditions, thereby satisfying the same-distribution assumption [14,15]. In real industrial environments, this assumption rarely holds due to highly dynamic and variable working conditions, ultimately leading to a significant decline in model generalization performance across operating conditions.

Corresponding author: Changqing Shen (e-mail: cqshen@suda.edu.cn).

To alleviate data distribution discrepancies caused by varying operating conditions, researchers have begun to explore how to endow diagnostic models with cross-domain adaptation (DA) capabilities, thereby enhancing their robustness under different working environments [16]. Within this context, the transfer learning (TL) paradigm has gradually emerged as an effective approach to addressing operating-condition variations and has been extensively studied and applied in the field of fault diagnosis [17,18]. Among TL approaches, DA methods aim to exploit limited target-domain samples to reduce the discrepancy between the source and target domains through distribution alignment and feature transfer, ultimately improving cross-domain diagnostic performance [19]. For example, Wang *et al.* [20] proposed a dynamic collaborative adversarial DA network for unlabeled cross-domain rotating machinery fault diagnosis. Qin *et al.* [21] developed a discriminative popular DA framework to mitigate classifier degradation caused by overlapping fault types across domains.

However, DA methods typically assume the availability of partial target-domain samples during training—an assumption that is often impractical in many real-world industrial scenarios. To address this limitation, researchers further introduced domain generalization (DG) methods to enhance a model's cross-domain generalizability in the absence of target-domain data [22,23]. For instance, Song *et al.* [24] proposed a contrast-assisted domain-specific removal network that extracts transferable representations by eliminating nontransferable information from domain-specific features, thereby enabling cross-domain fault diagnosis. Li *et al.* [25] integrated a domain expansion module with a dual-branch feature fusion network to improve rotating machinery fault identification under varying operating conditions.

Fault diagnosis methods based on the TL paradigm have achieved remarkable progress in addressing cross-condition and cross-equipment distribution discrepancies. By leveraging feature alignment, distribution matching, and the learning of domain-invariant representations, these methods substantially enhance model generalization under static cross-domain scenarios. However, they are typically built upon static datasets and commonly assume that the relationship between the source and target domains remains fixed during training. In reality, the operating environments of industrial equipment are inherently dynamic and continuous. The underlying data distributions evolve over time, operating conditions, and equipment states, leading to significant performance degradation when models are exposed to continuously arriving data from new domains.

Unlike traditional methods built on static datasets, incremental learning enables models to be dynamically updated as new data continuously arrive, thereby maintaining adaptability while effectively mitigating catastrophic forgetting. In recent years, various incremental learning-based studies have been applied to rotating machinery fault diagnosis. For example, Xu *et al.* [26] developed an incremental, broad learning method based on Dempster–Shafer (DS) theory for scenarios involving gradually increasing fault types, achieving continuous recognition of newly introduced types. Shen *et al.* [27] proposed an incremental learning method incorporating a dynamic branch-layer fusion mechanism, which significantly improved diagnostic performance under imbalanced dataset conditions. Tao *et al.* [28] designed a semi-supervised, contrastive-feature incremental learning framework that enables incremental fault diagnosis of rotating machinery with limited labeled samples. These approaches effectively alleviate catastrophic forgetting across various class-incremental scenarios, demonstrating both the potential and broad applicability of the incremental learning paradigm in fault diagnosis. However, as illustrated in Fig. 1, when substantial domain discrepancies exist between old and newly arriving data in incremental stages, the model often struggles to maintain a stable cross-domain feature space. The clustering boundaries of fault types from earlier domains are prone to drift, leading to marked performance degradation. Moreover, many existing methods heavily depend on limited incremental samples and are thus highly sensitive to sample quality, resulting in insufficient adaptability in domain-incremental scenarios.

To address the issue of insufficient model adaptability to new fault domains in the absence of old data, this study proposes a dual-component elastic adaptive network (DCEAN) to enhance incremental fault diagnosis of rotating machinery under varying operating conditions. Specifically, a sensitive parameter constraint (SPC) mechanism is first designed to suppress excessive updates of critical parameters based on their importance weights, thereby preserving historical knowledge at the parameter level. Next, by explicitly modeling feature drift, a feature drift self-calibration (FDSC) mechanism is developed to estimate and compensate for distributional shifts induced by operating-condition variations, thus maintaining cross-condition feature consistency. Experiments on a variable-condition bearing dataset demonstrate that the proposed DCEAN significantly outperforms existing methods in new-domain fault identification tasks and exhibits improved robustness and generalization. Ablation studies further validate the effectiveness of each component. The major contributions are summarized as follows:

1) A DCEAN framework without revisiting old samples is proposed to address the challenge of continual incremental fault diagnosis for rotating machinery under diverse and evolving operating conditions. Through dual-component collaborative optimization, the model effectively enhances domain adaptability and diagnostic stability.
2) At the parameter level, an SPC mechanism is introduced to adaptively preserve historical knowledge. This module evaluates the relative importance of different adapters through a channel-weighted optimization strategy and incorporates output perturbation analysis to capture the contribution of local adaptive features, thereby achieving a dynamic balance between model plasticity and knowledge stability.

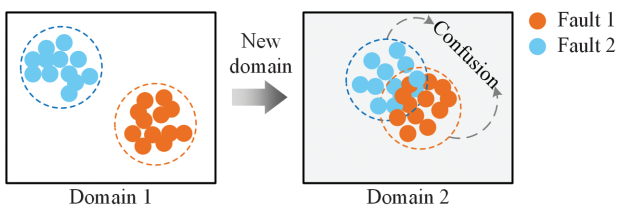


Fig. 1. Feature drift of old fault types in the new domain.

3) At the feature level, an FDSC mechanism is constructed to suppress and compensate prototype shifts caused by operating-condition drift. During training, a regularization constraint is employed to reduce deviations between the estimated feature drift and the null space, ensuring the continuity and robustness of decision boundaries and maintaining classifier consistency in new-domain tasks.

The remainder of this paper is organized as follows: Section II presents the relevant theoretical background; Section III provides a detailed description of the proposed DCEAN framework; Section IV reports the experimental validation; and Section V concludes the paper and discusses future research directions.

II. RELATED WORKS

A. INCREMENTAL LEARNING

Incremental learning, as a learning paradigm designed for dynamic data streams, has been extensively studied in fields such as computer vision and natural language processing [29]. Its core objective is to enable the model to accumulate knowledge and continuously adapt to new tasks or new classes without retraining from scratch.

From a formal perspective, incremental learning can be described as an optimization process conducted over a sequence of learning phases $\{1,2,\ldots,T\}$. The dataset at phase t is defined as follows:

$$D_t = \{(x_i^t, y_i^t)\}_{i=1}^{N_t}, \tag{1}$$

where x_i^t denotes the input samples, y_i^t represents their corresponding labels, and N_t is the number of samples. The goal at each phase is to learn a parameterized function f_{θ_t} that minimizes the empirical risk under the new data distribution $P_t(x,y)$, as follows:

$$\mathcal{L}_t(\theta_t) = \mathbb{E}_{(x,y)\sim D_t}[\ell(f_{\theta_t}(x), y)], \tag{2}$$

while retaining the knowledge acquired during the previous t-1 phases. Existing incremental learning methods can be broadly categorized into three types:

1) Replay-based methods: These approaches preserve part of the old samples or their feature representations for rehearsal to maintain the distribution of previously acquired knowledge [30]. Typically, a memory buffer $\mathcal{M}$ is constructed, and at each update step, the model minimizes a joint loss comprising current-task and replayed samples, as follows:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_t(\theta_t) + \lambda \mathcal{L}_{\text{replay}}(\theta_t; \mathcal{M}), \tag{3}$$

where λ is the balancing coefficient.

2) Regularization-based methods: These approaches introduce parameter-constraint terms into the loss function to prevent drastic parameter updates on new phases, thereby mitigating catastrophic forgetting.

3) Structure-expansion methods: These methods dynamically expand the network architecture (e.g., by adding new subnetworks or parameter modules) to acquire new knowledge. Such approaches maintain performance on previous phases while effectively isolating interference between different phases.

In summary, the essence of incremental learning lies in achieving a stability–plasticity balance, that is, maintaining the stability of previously acquired knowledge while enabling sufficient plasticity for learning new knowledge. However, standard incremental learning typically assumes that distributional changes arise solely from phase or fault-type increments, while insufficiently considering variations within the data domain itself—a factor that is particularly critical in complex industrial systems.

B. DOMAIN INCREMENTAL LEARNING AND FEATURE DRIFT

Domain incremental learning (DIL) is an important paradigm within continual learning, and its core objective is to enable models to adapt to learning scenarios in which the data distribution evolves sequentially over time. Under this setting, both the input feature distribution and the conditional distribution may change as time progresses.

In DIL, the model can only access the training set of the current phase, D_τ^{train}, while the aggregated test set $D_{1:\tau}^{\text{test}}$ is used to evaluate the trained model. The aggregated test set $D_{1:\tau}^{\text{test}}$ is defined as follows:

$$D_{1:\tau}^{test} = D_1^{test} \bigcup D_2^{test} \bigcup \cdots \bigcup D_\tau^{test}. \tag{4}$$

The model must maintain cross-DG while avoiding forgetting knowledge from previous domains when adapting to new ones. During the domain-incremental process, the change in the model's feature distribution can be formalized as follows:

$$\Delta f = \left\| \mu_t - \mu_{t-1} \right\|_2, \tag{5}$$

where $\mu_t = \mathbb{E}_{x\sim D_t}[f_{\theta_t}(x)]$ denotes the mean feature representation at phase t. A large Δf indicates that the feature space learned in the new domain deviates significantly from that of the previous domain, i.e., feature drift.

III. PROPOSED METHOD

A. PROBLEM DEFINITION AND OVERALL FRAMEWORK

In practical industrial systems, rotating machinery typically operates over long periods under continuously varying working conditions, causing the distribution of monitoring data to drift with changes in time and environment. When the operating condition shifts, traditional DL methods often fail because their learned feature representations and decision boundaries become invalid. Moreover, due to the large number of model parameters, directly fine-tuning the model on a new working condition leads to catastrophic forgetting, i.e., the loss of knowledge learned from previous conditions.

To ensure stable diagnostic performance across varying conditions, this study formulates the cross-condition learning problem of rotating machinery as a DIL task. The objective of DIL is to learn a feature extractor $F(x)$ and a classifier $g(\cdot)$ such that, under a new operating condition τ, the model maintains sufficient stability to preserve knowledge from previous conditions while retaining the plasticity needed to incorporate new condition-specific features.

The proposed overall structure of DCEAN is illustrated in Fig. 2. The feature extractor $F(x)$ consists of an adapter A_τ integrated with a pretrained transformer. The adapter A_τ

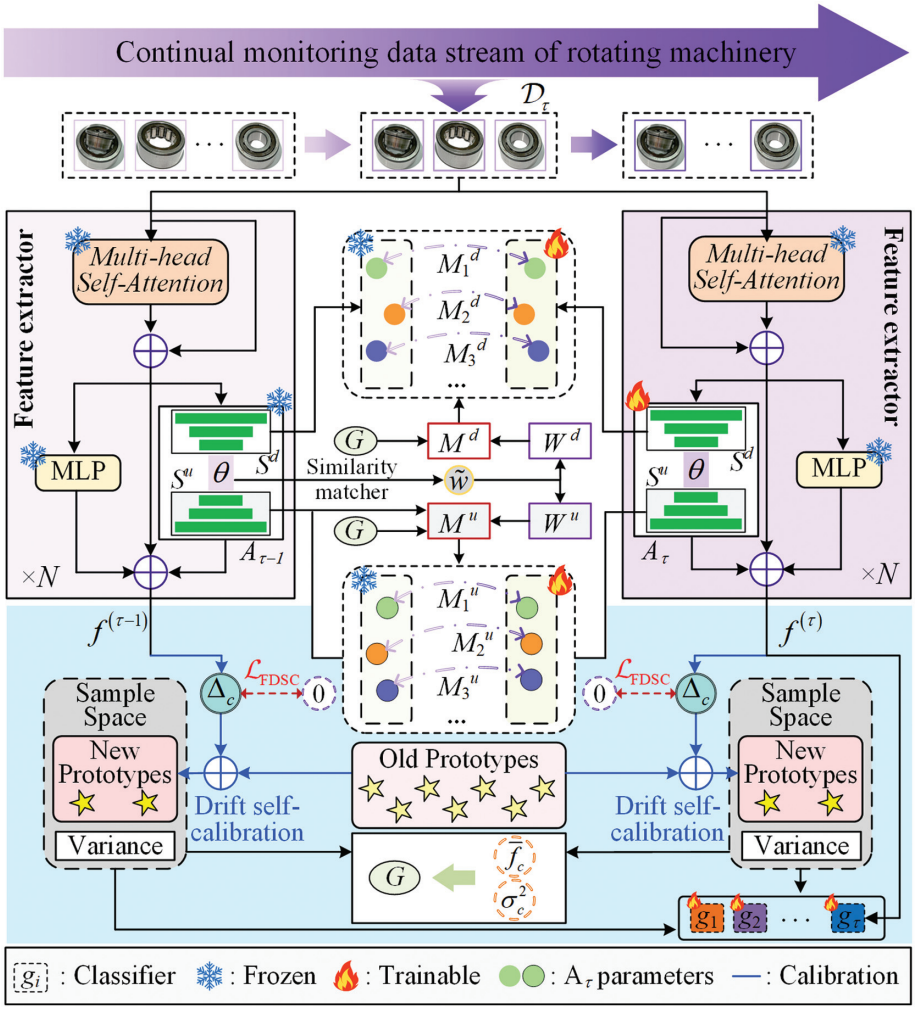


Fig. 2. The proposed overall framework for DCEAN.

adopts an encoder–decoder architecture, which typically includes a downsampling MLP layer S^d, an activation function $\theta(\cdot)$, and an upsampling MLP layer S^u. For an input x, the output is defined as follows:

$$y = \theta(x \cdot S^d) \cdot S^u = w \cdot S^u, \tag{6}$$

where w denotes the intermediate representation within the adapter. In addition, DCEAN is primarily composed of two key modules: (1) SPC, which constrains disruptive updates to critical parameters at the parameter level, thereby achieving robust preservation of previously learned knowledge; (2) FDSC, which compensates for distributional shifts caused by changing operating conditions at the feature level, ensuring semantic consistency across conditions. Here, elastic denotes an importance-aware adaptation mechanism that selectively adjusts parameter updates and feature representations to balance knowledge preservation and adaptability under operating-condition shifts.

B. SENSITIVE PARAMETER CONSTRAINT

To prevent the model from forgetting previously acquired knowledge during training on a new operating phase, the SPC module imposes differentiated constraints on the updates of critical parameters by evaluating their importance to past knowledge. The more sensitive a parameter is, the greater its impact on the performance of the original task when altered, and thus it should be subject to stricter constraints.

First, let the downsampling and upsampling parameters of the adapter at layer l under the $\tau-$th task be denoted as $\Theta^d_{l,\tau}$ and $\Theta^u_{l,\tau}$, respectively. The parameter-constraint loss is defined as follows:

$$\mathcal{L}_{\text{SPC}}^{(\tau)} = \sum_{l=1}^{L} \langle M^d_{l,\tau-1}, (\Theta^d_{l,\tau-1} - \Theta^d_{l,\tau})^2 \rangle + \sum_{l=1}^{L} \langle M^u_{l,\tau-1}, ((\Theta^u_{l,\tau-1})^\top - (\Theta^u_{l,\tau})^\top)^2 \rangle, \tag{7}$$

where $M^d_{l,\tau-1}$ and $M^u_{l,\tau-1}$ represent the importance matrices of the downsampling and upsampling parameters from the previous phase, and L denotes the number of layers containing trainable adapters. Subsequently, the global importance is obtained through the statistics of the mean and variance of each fault type, as follows:

$$G_\tau = G_{\tau-1} + \frac{1}{N_c^{(\tau-1)}} \sum_{c \in C} \frac{\bar{f}_c}{\sigma_c^2}, \tag{8}$$

where $\bar{f}_c$ and σ_c^2 denote the mean and variance of the features corresponding to fault type c, and $N_c^{(\tau-1)}$ is the number of fault types in the previous phase.

Cross-domain samples often exhibit substantial discrepancies in local feature distributions. To avoid feature degradation caused by such distributional mismatch, we introduce a similarity-based feature matcher. Specifically, taking the global feature unit $w_{l,\tau}[0,:]$ as a reference, each local feature unit $w_{l,\tau}[j,:]$ is aggregated through cosine-similarity–based weighting, producing a more robust and domain-consistent representation $\tilde{w}_{l,\tau}$. The global-aligned feature $\tilde{w}_{l,\tau}$ and the local sensitivity $W^d_{l,\tau}$ of the downsampling matrix in layer l are defined as follows:

$$\begin{cases} \tilde{w}_{l,\tau} = \sum_{j=0}^{d_t} \left\langle \frac{w_{l,\tau}[j,:]}{\|w_{l,\tau}[j,:]\|}, \frac{w_{l,\tau}[0,:]}{\|w_{l,\tau}[0,:]\|} \right\rangle w_{l,t}[j,:] \\ W^d_{l,\tau} = W^d_{l,\tau-1} + \frac{1}{N^{(\tau)}} \sum_{i=1}^{N_\tau} \tilde{w}_{l,\tau}(x_i) \end{cases}, \tag{9}$$

where d_t is the length of the feature sequence for the input sample. The local sensitivity $W^u_{l,\tau}$ of the upsampling matrix at layer l is defined as:

$$\begin{cases} w_{l,\tau} = \tilde{w}_{l,\tau} \cdot \sum_{k=1}^{m} |q_{l,\tau}[:,k]| \\ W^u_{l,\tau} = W^u_{l,\tau-1} + \frac{1}{N^{(\tau)}} \sum_{i=1}^{N_\tau} w_{l,\tau}(x_i) \end{cases}, \tag{10}$$

where $q_{l,t}[:,\mathrm{k}]$ denotes the absolute weight value in the upsampling matrix that flows from the intermediate feature unit to the k-th output unit, and d is the feature dimension.

Global importance G_τ aims to quantify how critical each adapter parameter is to maintaining the overall discriminative structure of previously learned fault types. From a statistical perspective, parameters that contribute to stable and well-separated class representations across historical phases should be preserved more strictly. Therefore, global importance is estimated at the class level by aggregating the mean and variance of feature representations associated with each fault type learned in the previous phase. This measure reflects the long-term and global contribution of adapter parameters to historical knowledge retention, independent of individual samples.

In contrast, local sensitivity $W_{l,\tau}$ focuses on the influence of adapter parameters on the propagation of local intermediate features to the model output. During incremental learning, even small perturbations in certain parameters may cause disproportionate changes in feature responses, especially under domain shifts. To capture this behavior, local sensitivity is evaluated by analyzing the contribution of intermediate adapter features to the output through similarity-weighted aggregation. Parameters that strongly affect local feature alignment and output stability are regarded as highly sensitive at the sample and feature levels. Finally, by integrating G_τ and $W_{l,\tau}$, the overall parameter importance is computed as:

$$\begin{cases} M^u_{l,\tau} = \beta_1 G_\tau^\top \odot W^u_{l,\tau} \\ M^d_{l,\tau} = \beta_2 G_\tau^\top \odot W^d_{l,\tau} \end{cases}, \tag{11}$$

where β_1 and β_2 denote the amplification and scaling coefficients for global and local importance, respectively.

By jointly considering global importance and local sensitivity, SPC achieves a complementary assessment of parameter significance. Global importance emphasizes class-level stability and historical knowledge preservation, while local sensitivity captures feature-level responsiveness under new operating conditions. Their integration enables differentiated parameter constraints, allowing critical parameters to be protected while permitting sufficient flexibility for noncritical parameters during incremental adaptation.

C. FEATURE DRIFT SELF-CALIBRATION

SPC first evaluates parameter importance based on global statistics and local sensitivity, and then applies weighted constraints to the shared adapters accordingly, thereby balancing model plasticity with the preservation of discriminative structures from previous operating conditions. However, relying solely on parameter-level preservation may still cause decision shifts in the unified classifier when encountering cross-condition variations. To address this issue, the FDSC module is designed to explicitly model feature drift and maintain semantic consistency through a trainable self-calibration mechanism.

Let $F_{\tau-1}$ denote the feature extractor of the previous phase, and F_τ the one for the current phase. The sample-level drift is defined as:

$$\begin{cases} \delta_i^{(\tau)} = f_i^{(\tau-1)} - f_i^{(\tau)} \\ f_i^{(\tau)} = \mathrm{F}_\tau(x_i) \end{cases}, \tag{12}$$

where $f_i^{(\tau)}$ represents the features extracted by F_τ. To obtain a stable drift estimate, similarity-weighted aggregation is applied over samples, as follows:

$$\Delta_c^{(\tau-1\to\tau)} = \frac{\sum_i \alpha_i \delta_i^{(\tau)}}{\sum_i \alpha_i}, \tag{13}$$

$$\alpha_i = \exp\left(- \frac{\left\| f_i^{(\tau-1)} - p_c^{(\tau-1)} \right\|^2}{2\sigma_c^2} \right). \tag{14}$$

During training, a drift-regularization term is introduced to prevent noise amplification, as follows:

$$\mathcal{L}^{(\tau)}_{\mathrm{FDSC}} = \sum_{c\in\mathcal{C}} \left\| \Delta_c^{(\tau-1\to\tau)} \right\|_2^2. \tag{15}$$

The fault prototypes are subsequently updated as:

$$P_c^{(\tau)} = \begin{cases} p_c^{(r-1)} + \Delta_c^{(r-1\to\tau)}, & c \notin \mathcal{C}_{\mathrm{old}} \\ \frac{1}{n_c} \sum_{i=1}^{n_c} f_{i,c}^{(r)}, & c \in \mathcal{C}_{\mathrm{old}} \end{cases} \tag{16}$$

Here, a fault prototype is defined as the class-level centroid of feature embeddings for a given fault type, obtained by averaging the extracted features of samples belonging to that type. It serves as a compact semantic representation of fault features in the feature space. Through Eqs. (12)–(16), the FDSC module achieves adaptive alignment of feature semantics, thereby enhancing the robustness of cross-condition fault diagnosis.

D. APPLICATION OF DCEAN IN THE FAULT DIAGNOSIS OF ROTATING MACHINERY

DCEAN integrates the SPC and FDSC modules in a unified manner. When a new operating condition arrives, the model performs adaptive adjustments at both the parameter level and the feature level, enabling dynamic knowledge preservation and refinement. The overall training objective is formulated as follows:

$$\mathcal{L}^{(\tau)} = \underbrace{\mathcal{L}^{(\tau)}_{\mathrm{ce}}}_{\text{classification loss}} + \lambda_1 \mathcal{L}^{(\tau)}_{\mathrm{SPC}} + \lambda_2 \mathcal{L}^{(\tau)}_{\mathrm{FDSC}}, \tag{17}$$

Algorithm 1. DCEAN for domain-incremental fault diagnosis

Input: Sequential domain datasets $\{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_\tau\}$

Output: Updated model parameters A_τ and class prototypes $p_c^{(\tau)}$

Initialize model with $\mathcal{D}_1 \rightarrow$ train $F(x; A_\tau)$, $g(\cdot)$, obtain $\{p_c^{(1)}\}$
For $\tau = 2$ to T **do**
  Load previous parameters $A_{\tau-1}$ and prototypes $p_c^{(\tau-1)}$
  Compute importance matrices $M_{l,\tau-1}$ using Eqs. (8)–(11)
  For each batch in $\mathcal{D}_\tau$ **do**
    Extract features $f^\tau = F_\tau(x)$
    Estimate drift Δ_c using Eqs. (12)–(14)
    Update prototypes $p_c^{(\tau)}$ with Eq. (16)
    Compute the total loss $\mathcal{L}^{(\tau)}$ using Eq. (17).
    Backpropagate and update A_τ
  end for
  Save $\{A_\tau, p_c^{(\tau)}, M_{l,\tau-1}\}$ for next domain
End for

where λ_1 and λ_2 are weighting coefficients. During training, the backbone network remains frozen, while the shared adapters are updated under the constraints imposed by SPC, and the unified classifier is dynamically calibrated through FDSC. SPC alleviates catastrophic forgetting without introducing additional parameters. Meanwhile, FDSC estimates feature drift and introduces a regularization constraint to compensate for the semantic shift of old fault-type prototypes, ensuring consistent and stable discrimination in the unified classifier under continuously varying operating conditions. Through the synergy of these two mechanisms, DCEAN achieves a unified balance between elastic adaptation and knowledge retention in dynamic environments. Algorithm 1 outlines the entire training procedure of DCEAN.

IV. EXPERIMENTS

This section first presents the configuration of the experimental datasets, followed by the experimental settings and performance evaluation protocols. The proposed method is then systematically compared with current state-of-the-art approaches. In addition, ablation studies are designed to analyze the contribution of each key module to the overall performance, thereby validating the rationality and effectiveness of the model design.

A. DATASET DESCRIPTION

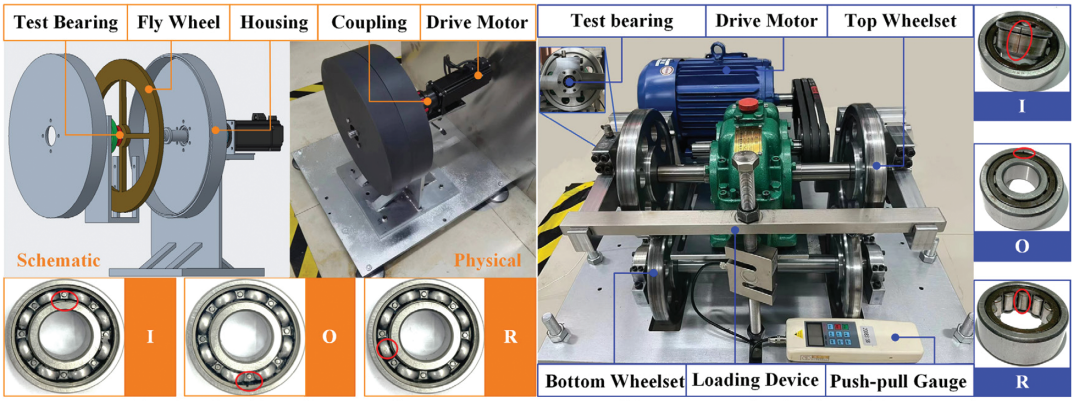
To comprehensively evaluate the generalization and robustness of the proposed model under continual domain-incremental phases, two representative bearing datasets are employed.

The flywheel bearing dataset is collected from a satellite flywheel bearing fault simulation platform, whose structural setup and samples from three typical states are illustrated in Fig. 3(a). The platform consists of a drive motor, a coupler, a flywheel, and a housing. Vibration signals are measured using acceleration sensors at rotational speeds of 600, 800, 1000, and 1200 rpm with a sampling rate of 10 kHz to simulate continuously varying operating conditions. Five state classes are included: healthy (H), inner-race fault (I), outer-race fault (O), rolling-element fault (R), and mixed outer-race and rolling-element fault (OR).

The wheelset bearing dataset is acquired from an integrated wheelset drive test rig, whose structure and three typical state samples are shown in Fig. 3(b). The platform is composed of a drive motor, a loading device, and upper–lower wheelsets to emulate the complex loading conditions during train operation. Wheelset bearing vibration signals are collected at 800 rpm under loads of 0, 0.8, 1.6, and 2.4 kN with a sampling rate of 12.8 kHz. The health state settings are consistent with those of the flywheel bearing dataset. The bearing types, operating conditions, and fault types involved in both datasets are summarized in Table I.

B. EXPERIMENTAL SETUP

The experimental setup covers a continual fault diagnosis scenario consisting of an initial phase and three incremental phases, with each phase containing five different fault types. For each fault type under each operating condition, 100 samples are used for training and 100 samples for testing. During incremental training, no samples from previously learned fault types are stored or reused. Instead, class prototypes are preserved to represent historical knowledge, enabling exemplar-free incremental learning while reducing storage requirements and addressing potential privacy concerns. A standard pretrained transformer encoder is adopted as the backbone network. The backbone



(a) Flywheel bearing test rig (b) Integrated wheelset drive test rig

Fig. 3. Flywheel and wheelset bearing test rig.

Table I. Detailed description of the flywheel and wheelset bearing datasets

Dataset	Bearing	Working condition	Status type	Fault diameter (mm)	Label	Phase
Flywheel	6203DDU NSK	600 rpm	H, I, O, R, OR	0.3	0–19	0
		800 rpm				1
		1000 rpm				2
		1200 rpm				3
Wheelset	NT204ET	0 kN		0.5		0
		0.8 kN				1
		1.6 kN				2
		2.4 kN				3

Table II. Experimental parameter settings

Parameter	β_1	β_2	λ_1	λ_2	lr	Batch size	Epochs
Value	100	1	10	10	0.01	32	100

architecture is kept fixed for all methods to ensure fair comparison, and no task-specific modifications are introduced. Performance gains are therefore mainly attributed to the proposed SPC and FDSC mechanisms rather than to backbone engineering. The parameter configurations of the proposed method are provided in Table II.

The experiments primarily adopt the following performance metrics to evaluate all models:

(1) Accuracy: This metric measures the overall classification performance of a model across all learned phases. It is computed as the average of the final test accuracies over all phases, formulated as:

$$\text{Accuracy} = \frac{1}{T}\sum_{t=1}^{T} a_{t,T} \times 100\%$$

where T denotes the total number of phase tasks, and $a_{t,T}$ represents the model's test accuracy on the t-th phase after completing learning on all T phases.

(2) F1-score (F1): This metric evaluates the comprehensive classification performance and mitigates misleading assessments caused by data imbalance, i.e., ensuring that precision and recall are jointly considered, as follows:

$$\text{F1} = 2 \times \text{Precision} \times \text{Recall}/(\text{Precision} + \text{Recall})$$

To investigate the effectiveness and superiority of the proposed DCEAN in continual fault diagnosis under continuously varying operating conditions, comparisons are conducted with several advanced methods. In addition, a series of ablation studies is performed to thoroughly analyze the contribution of each component within DCEAN. The methods and their details are summarized in Table III. Among them, Finetune represents a common strategy for TL. LwF [31], iCaRL [32], WA [33], and ADC [34] are representative incremental learning methods selected to

Table III. Detailed description of the comparison and ablation methods

Method	Description
Finetune	Fine-tunes all network parameters.
LwF	Performs incremental learning without old data by adding a penalty term in the loss to constrain changes in the classifier for previous phases.
iCaRL	Balances learning of old and new phases via nearest-neighbor classification combined with exemplar replay.
WA	Maintains discrimination of old fault types through knowledge distillation and corrects bias in the fully connected layer via weight alignment to ensure fairness between old and new classes.
ADC	Uses adversarial perturbations to pull current-sample embeddings toward old-class prototypes, estimating and compensating for feature drift to restore old-class performance in sample-free continual learning.
DCEAN	The proposed method.
w/o FDSC	Ablation model used to verify the effectiveness of the FDSC module.
w/o SPC	Ablation model used to verify the effectiveness of the SPC module.

Table IV. The diagnostic accuracy (%) of flywheel bearings under variable operating conditions

Method	Phase 0	Phase 1	Phase 2	Phase 3	F1
Finetune	99.83 ± 0.16	50.19 ± 3.78	35.67 ± 3.68	23.79 ± 4.88	13.62
LwF	99.83 ± 0.16	57.59 ± 4.55	38.07 ± 3.77	27.19 ± 4.33	18.00
iCaRL	99.83 ± 0.16	55.38 ± 2.79	39.73 ± 3.28	36.93 ± 2.40	26.88
WA	99.83 ± 0.16	77.16 ± 3.66	65.73 ± 3.73	51.23 ± 4.08	30.68
ADC	99.83 ± 0.16	79.87 ± 2.48	69.21 ± 3.59	63.24 ± 4.31	58.64
DCEAN	**99.83** ± 0.16	**98.01** ± 0.21	**95.59** ± 0.63	**93.05** ± 0.92	**92.79**

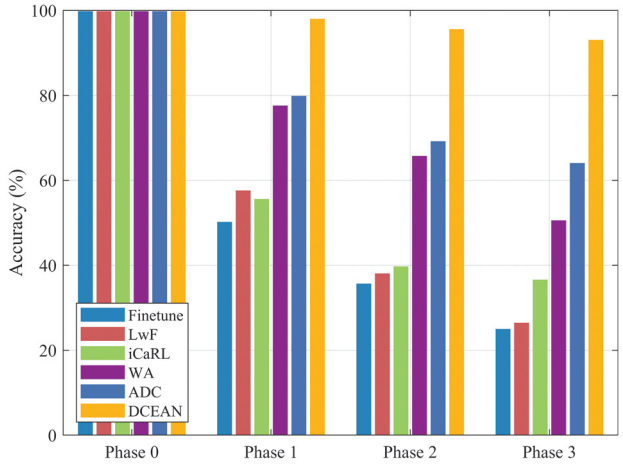


Fig. 4. Diagnostic accuracy of flywheel bearings at different phases.

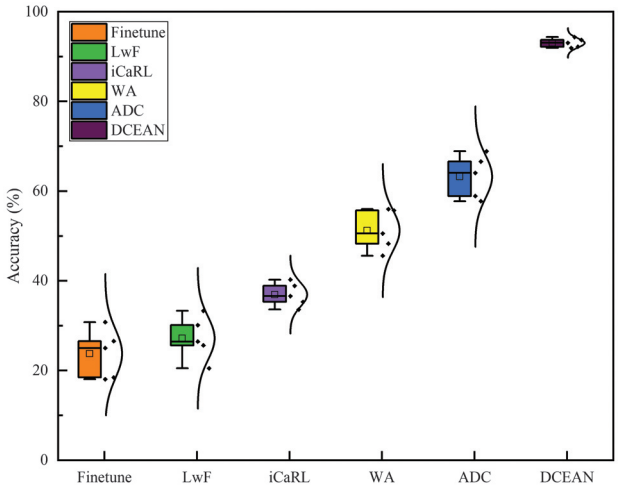


Fig. 5. Box-and-whisker plot versus scatter plot of accuracy rates for different models in Case 1.

cover both replay-based and exemplar-free paradigms, thereby enabling a comprehensive comparison under different incremental learning assumptions. Moreover, w/o SPC and w/o FDSC denote the ablation variants of DCEAN, used to verify the influence of each component on the proposed framework.

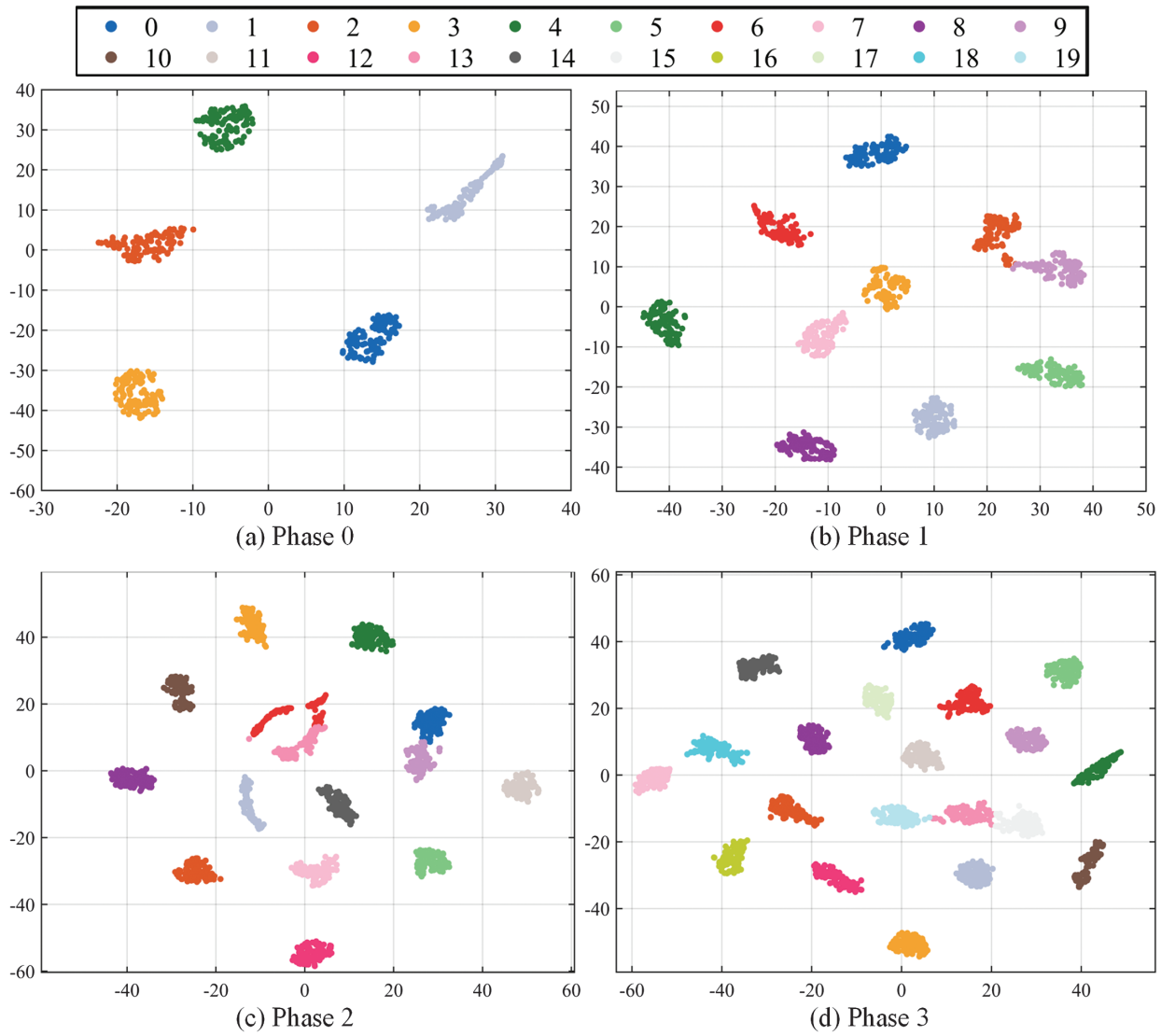


Fig. 6. Feature clustering of DCEAN at different phases in Case 1.

C. CASE STUDY

1). CASE 1: FLYWHEEL BEARINGS DATASET. This subsection presents the evaluation results of the proposed DCEAN model and several mainstream comparison methods on incremental fault diagnosis tasks using the flywheel-bearing dataset. To reduce random fluctuations in the experiments, each method was independently repeated five times, and the average results were reported as the final performance indicators. Table IV summarizes the diagnostic accuracy at each task phase as well as the F1 score at the final phase, providing quantitative evidence for overall performance comparison. As shown in Fig. 4, with the gradual increase of task phases, the diagnostic accuracy of traditional comparison methods exhibits a continuous downward trend, whereas DCEAN consistently maintains accuracy above 90%, demonstrating remarkable stability and superiority.

Figure 5 further illustrates the performance characteristics of each method across multi-phase tasks from a statistical distribution perspective. The combination of boxplots and scatter plots intuitively reflects the central

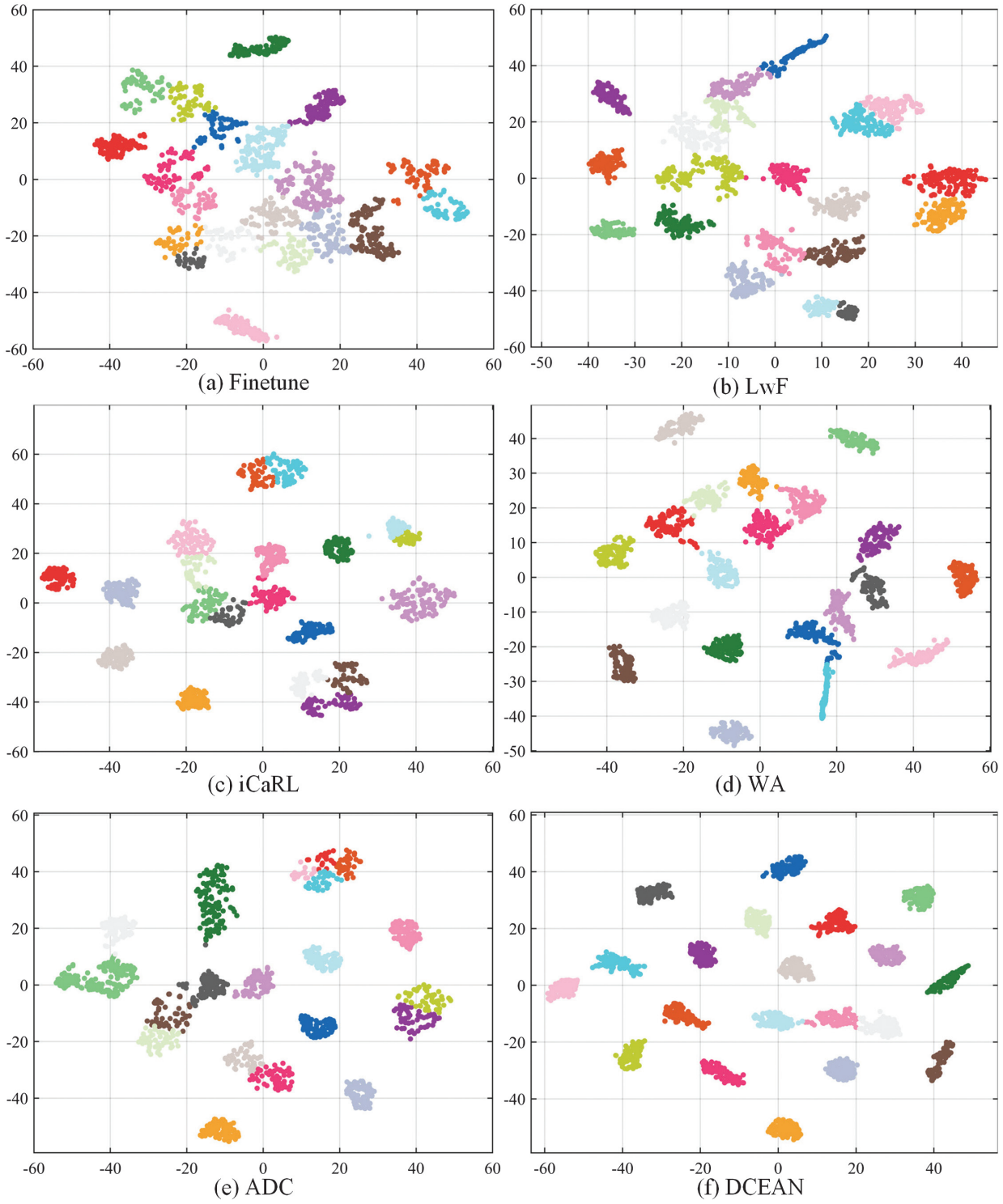


Fig. 7. Feature clustering results for different methods in the final testing phase of Case 1.

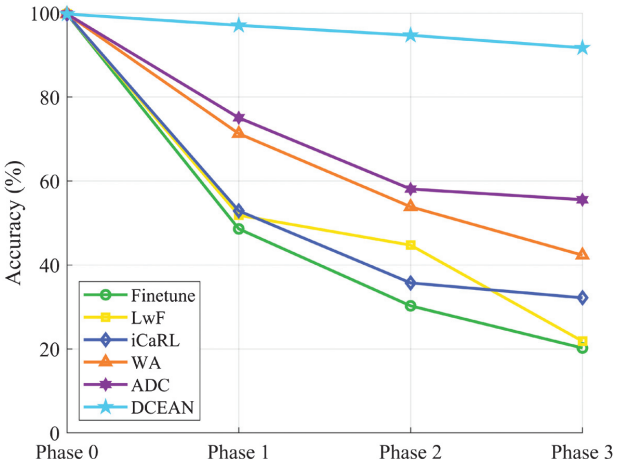


Fig. 8. Diagnostic accuracy of wheelset bearings at different phases.

tendency and dispersion of the prediction performance. The results show that DCEAN exhibits the most concentrated accuracy distribution, with compact box heights and very few outliers. This indicates that the method not only maintains high diagnostic performance at all phases but also shows extremely low performance variance, confirming its robustness and consistency in continual learning scenarios. In contrast, ADC and WA rank second, with overall accuracy superior to other baselines but still showing noticeable fluctuations. iCaRL, LwF, and Finetune exhibit the weakest performance, with significantly lower mean accuracy, clearly revealing their susceptibility to catastrophic forgetting in multi-phase incremental learning, making it difficult to effectively preserve historical knowledge.

To further validate the feature-learning capability of the proposed DCEAN model in cross-domain continual incremental tasks, this study employs t-SNE to visualize the feature representations at each phase, as shown in Fig. 6. From the visualizations, DCEAN is observed to form well-defined feature clusters at every phase, with high intra-class compactness and inter-class separability across different fault types. Only a few fault types show slight local boundary contact. Notably, even as the task scale expands from the initial 5 fault types to 20 fault types, the model consistently maintains strong discriminative feature representation. This indirectly demonstrates that DCEAN effectively prevents forgetting of prior knowledge and enables continuous adaptation to newly introduced domains and evolving data distributions in cross-DIL.

To further compare the differences in feature learning between methods, Fig. 7 presents the t-SNE distributions of DCEAN and other baselines at the final testing phase. As shown in Fig. 8(a)–(e), most traditional methods exhibit varying degrees of feature overlap after multi-phase training, with blurred class boundaries and loose cluster structures, indicating their difficulty in maintaining discriminative power for both old and new classes during incremental learning. In contrast, the feature distribution of DCEAN retains clear and stable cluster patterns, with well-separated class boundaries and almost no obvious overlapping. These results demonstrate the method's strong capability in preserving the integrity of the feature representation space during continual cross-domain learning, further supporting its advantages in mitigating catastrophic forgetting and addressing feature drift.

2). CASE 2: WHEELSET BEARINGS DATASET. Table V presents the multi-phase fault diagnosis accuracy of all methods on the wheelset bearing dataset. It can be observed that all methods achieve nearly identical high accuracy in the initial phase, indicating that different models are capable of robust feature learning when the data distribution remains unchanged. However, as the phases progress, the performance of conventional methods deteriorates rapidly, revealing severe catastrophic forgetting. For example, the accuracy of the Finetune method drops sharply after Phase 1 and retains only about 20% recognition capability by Phase 3. Although LwF and iCaRL show slight improvements, they still fail to maintain stable cross-phase diagnostic performance. WA and ADC achieve relatively better results, yet their accuracy continues to decline as phases increase, reflecting persistent issues of knowledge loss and feature drift under domain-incremental scenarios.

In contrast, the proposed DCEAN model consistently maintains a significantly superior diagnostic performance across all phases. This result clearly demonstrates that DCEAN can effectively mitigate forgetting and robustly integrate new knowledge from varying operating conditions during cross-domain continual learning. The line chart shown in Fig. 8 further illustrates this trend: while the accuracy curves of all comparison methods drop substantially across phases, DCEAN consistently remains in a high-accuracy region with a smooth and stable trajectory.

Figure 9 shows the statistical distribution of accuracy for all methods across multi-phase tasks. It can be observed that the box plot of DCEAN is highly compact, with sample points displaying strong concentration and very few outliers, indicating excellent consistency and robustness across different experimental runs. In comparison, the box plots of other methods are generally more dispersed, demonstrating noticeable fluctuations and instability, once again confirming the superiority of DCEAN in domain-incremental learning.

Figure 10 illustrates the confusion matrices of DCEAN across all phases. It is evident that the method continually maintains strong discriminative capability for various fault

Table V. The diagnostic accuracy (%) of wheelset bearings under variable operating conditions

Method	Phase 0	Phase 1	Phase 2	Phase 3	F1
Finetune	99.60 ± 0.34	48.60 ± 3.03	30.27 ± 2.99	20.56 ± 3.30	11.81
LwF	99.60 ± 0.34	51.85 ± 2.81	44.73 ± 3.55	22.07 ± 4.20	15.71
iCaRL	99.60 ± 0.34	52.89 ± 2.76	35.73 ± 3.23	33.24 ± 3.76	26.37
WA	99.60 ± 0.34	71.30 ± 3.33	53.87 ± 3.01	43.26 ± 3.51	22.51
ADC	99.60 ± 0.34	75.07 ± 2.69	58.10 ± 2.92	55.23 ± 2.87	46.52
DCEAN	**99.60** ± 0.34	**97.10** ± 0.76	**94.73** ± 0.89	**91.65** ± 1.16	**90.89**

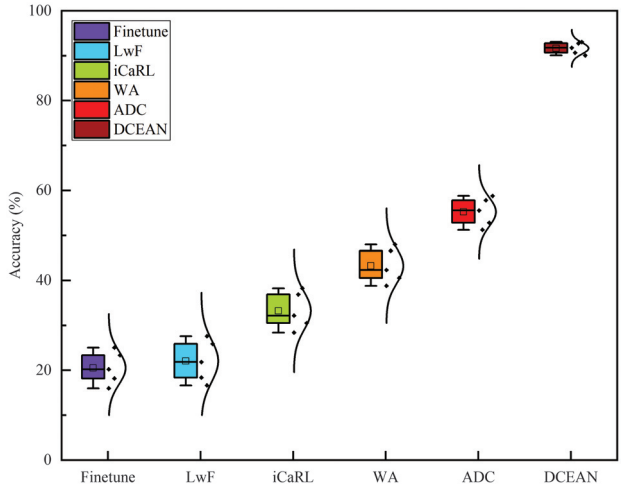


Fig. 9. Box-and-whisker plot versus scatter plot of accuracy rates for different models in Case 2.

types as the task evolves, with minimal inter-class confusion. This highlights its advantages in cross-domain knowledge integration and feature stability. To further compare the recognition capability of different methods in the final testing phase, Fig. 11 presents the confusion matrices of all methods. As shown in Fig. 11(a), Finetune is only able to correctly recognize the newly introduced fault type in the final operating condition and almost entirely loses knowledge from earlier phases, exhibiting a typical complete-forgetting phenomenon. Fig. 11(b)–(e) shows that although LwF, iCaRL, WA, and ADC retain partial historical information, their predictions still suffer from varying degrees of class confusion and knowledge loss, making it difficult to sustain stable fault diagnosis performance under domain-incremental scenarios. In contrast, the proposed DCEAN model in Fig. 11(f) demonstrates clear and balanced recognition across all classes. It not only preserves knowledge from earlier phases but also continues to learn effectively in new phases, achieving a coherent integration of old and new knowledge.

3). ABLATION STUDY. To evaluate the practical contributions of the two core modules—SPC and FDSC—in cross-condition continual incremental learning, comprehensive ablation studies were conducted on both datasets, with the results summarized in Table VI. Overall, the complete DCEAN model achieves the best fault diagnosis performance across all phases, and its accuracy exhibits the smallest degradation as phases progress. This further demonstrates that the synergistic design of SPC and FDSC is essential for maintaining cross-domain stability.

From the results of w/o FDSC, it can be observed that the model experiences a pronounced performance drop after

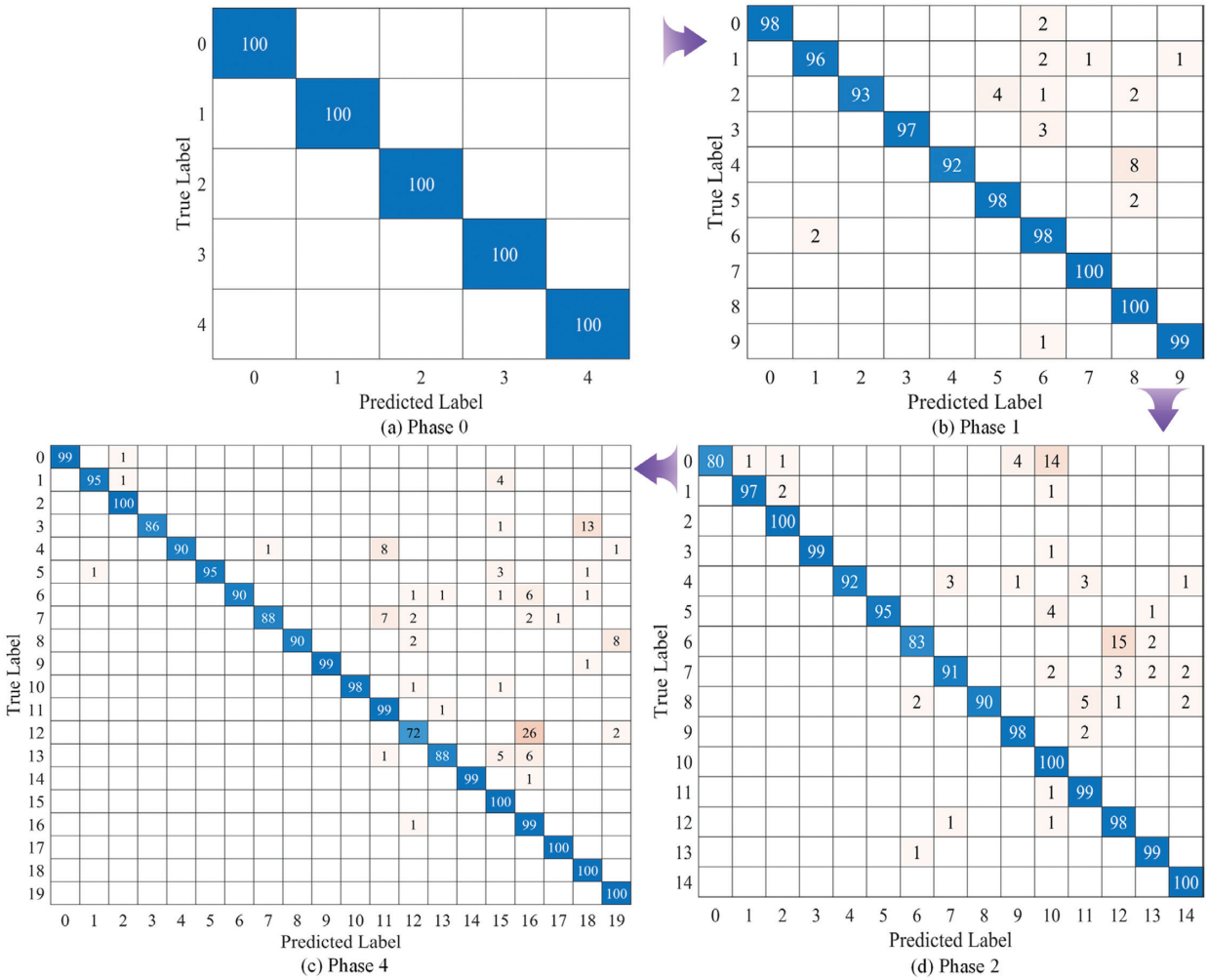


Fig. 10. Confusion matrices of DCEAN at different phases in Case 2.

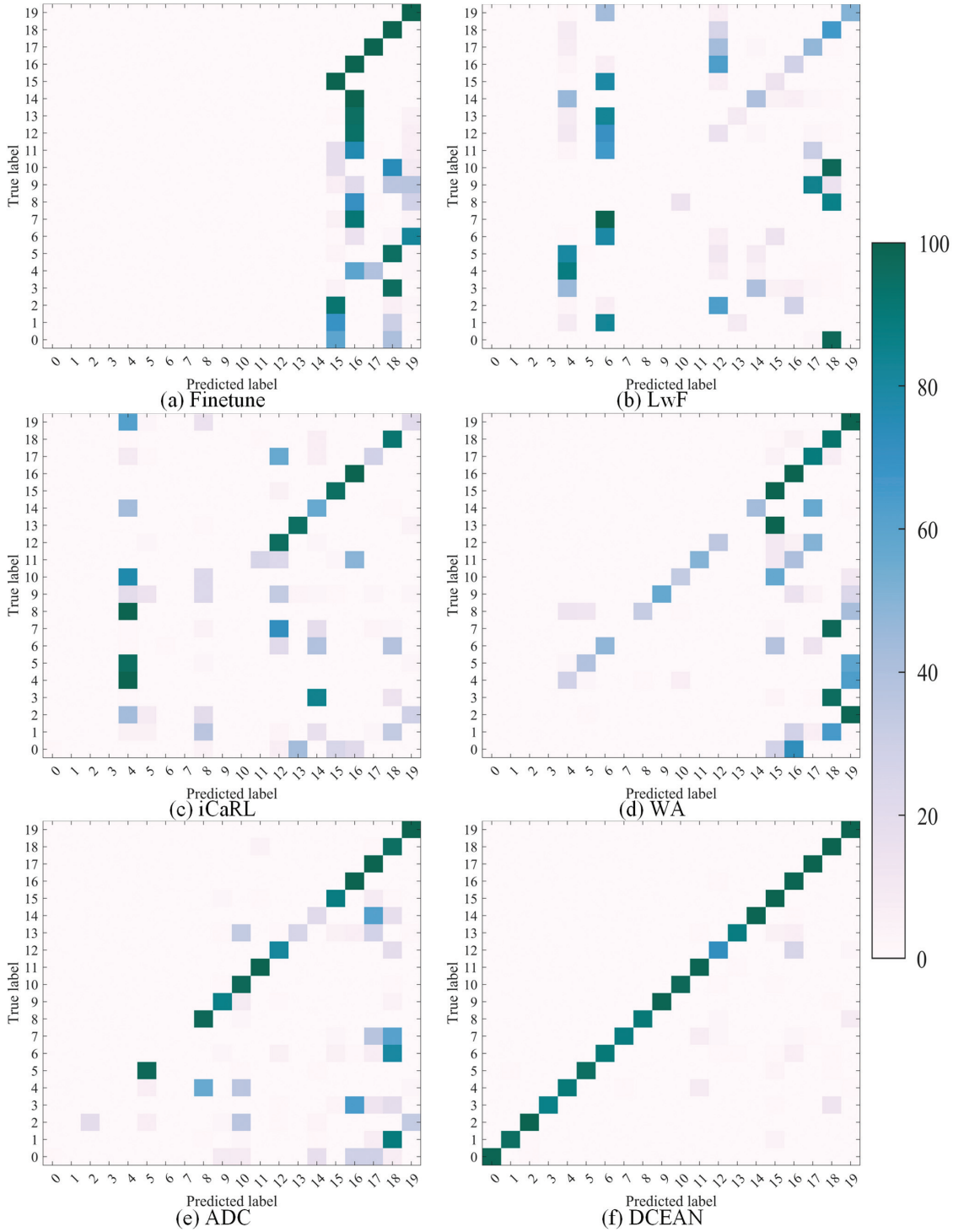


Fig. 11. Confusion matrices for different methods in the final testing phase of Case 2.

Table VI. The diagnostic accuracy (%) of the ablation methods under variable operating conditions

Method	Dataset	Phase 0	Phase 1	Phase 2	Phase 3
DCEAN	Flywheel	99.83	98.01	95.59	93.05
	Wheelset	99.60	97.10	94.73	91.65
w/o FDSC	Flywheel	99.83	96.35	87.21	84.81
	Wheelset	99.60	91.30	84.47	82.70
w/o SPC	Flywheel	99.83	97.59	89.73	88.60
	Wheelset	99.60	95.48	85.87	85.10

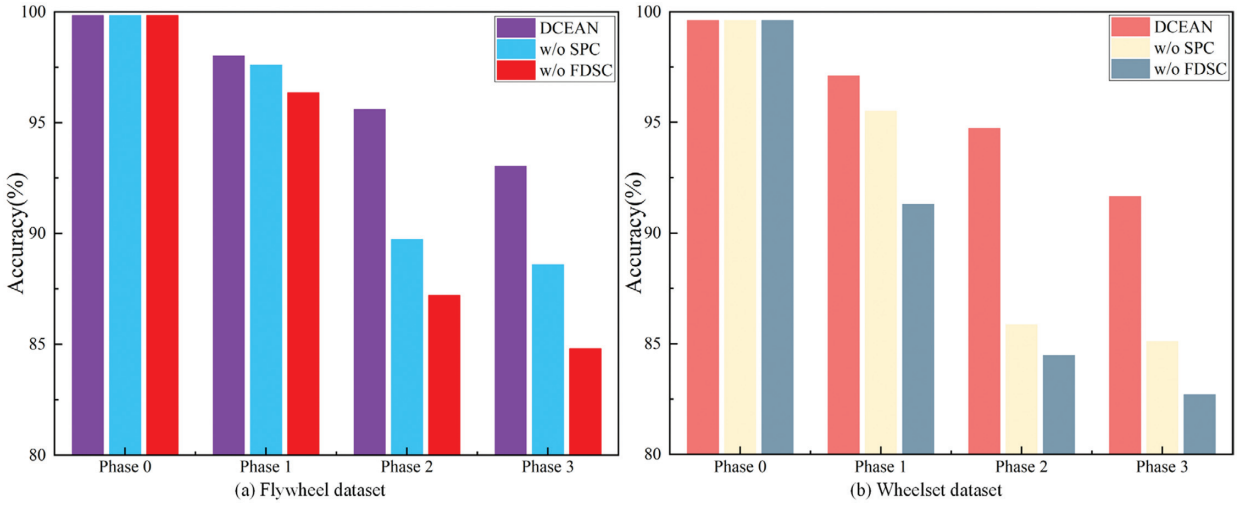


Fig. 12. Diagnostic accuracy of ablation methods at different phases.

Phase 1 on both the flywheel and wheelset bearing datasets. Without explicit feature-drift calibration, the model struggles to maintain stable semantic representations across different operating conditions. This inconsistency causes the shared classifier to shift its decision boundaries between old and new feature distributions, thereby intensifying

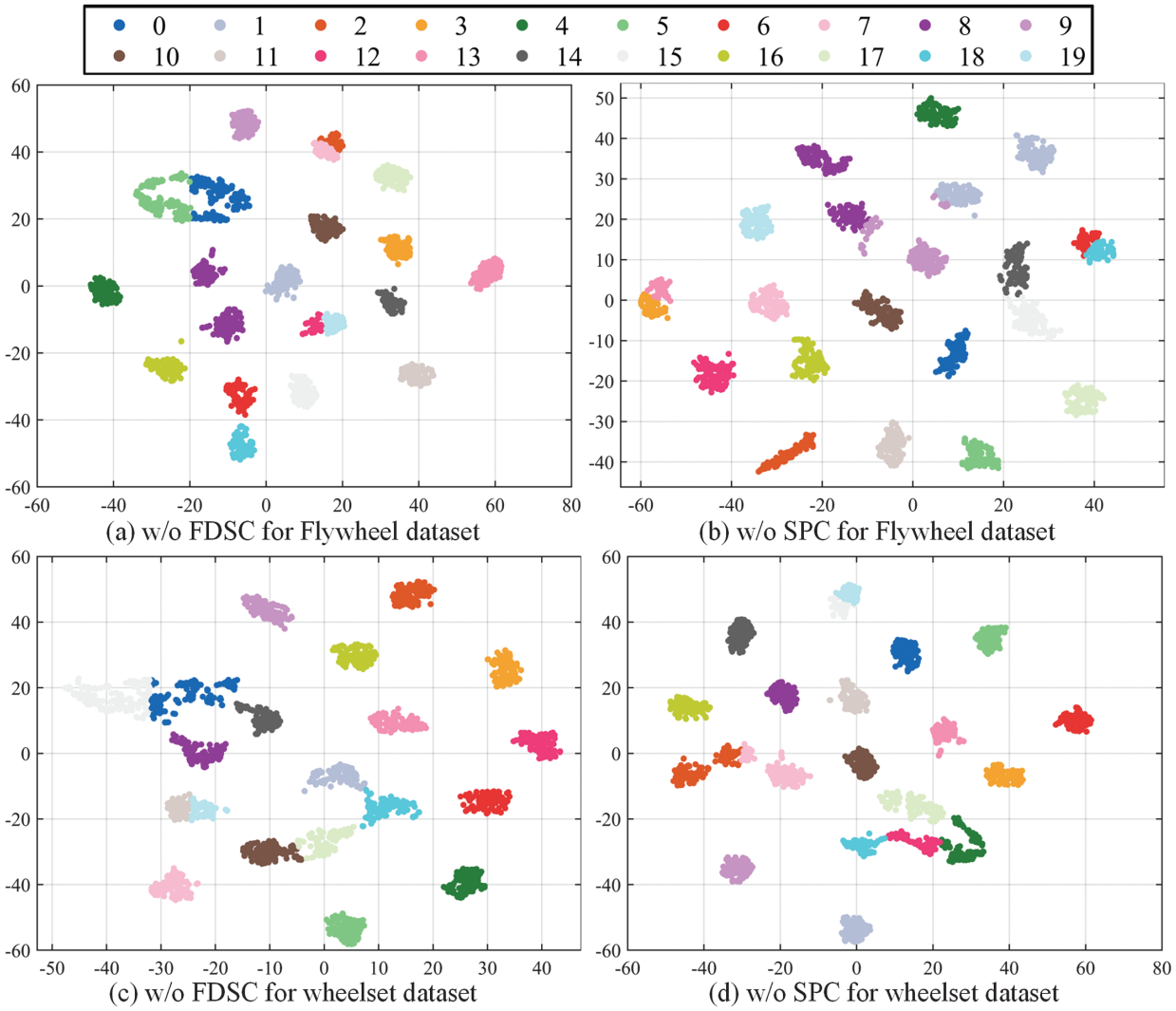


Fig. 13. Feature clustering of ablation methods in the final testing phase.

performance degradation. In contrast, the w/o SPC results indicate that although the model retains some cross-condition adaptability, the absence of parameter-importance constraints—based on statistical characteristics and sensitivity analysis—leads to excessive updates of critical parameters in new phases. As a result, discriminative structures learned from earlier conditions become weakened, allowing errors to accumulate and propagate into subsequent phases.

The bar charts in Fig. 12(a) and Fig. 12(b) further validate these findings: the complete DCEAN consistently maintains a significantly higher performance curve with a smoother downward trend, whereas both w/o FDSC and w/o SPC display substantially sharper degradation across phases. This clearly indicates that both modules play irreplaceable roles in sustaining continual learning performance. To analyze the module contributions from a feature-space perspective, Fig. 13 presents the t-SNE visualization results in the final testing phase. As shown in Fig. 13(a)–(d), removing either FDSC or SPC results in varying degrees of cluster overlap among different fault types, with the feature mixing in w/o FDSC being the most severe. This observation aligns fully with the numerical results: the absence of feature-drift calibration leads to semantic inconsistency across operating conditions, producing blurred classification boundaries; meanwhile, the lack of parameter-importance constraints partially disrupts the representational structure of previously learned knowledge, degrading feature clustering quality.

In summary, the ablation study demonstrates that SPC and FDSC contribute complementary functionalities in terms of parameter stability and feature semantic consistency. Their synergy constitutes the key mechanism enabling DCEAN to maintain high accuracy and stability in cross-condition incremental learning.

V. CONCLUSION

To address the challenges of incremental fault diagnosis for rotating machinery under continuously varying operating conditions, this study proposes a DCEAN with strong cross-condition adaptability. The method jointly models parameter stability and feature consistency by introducing a sensitive-parameter constraint mechanism and a feature-drift self-calibration mechanism, enabling knowledge preservation and cross-phase semantic alignment without relying on past samples. The former suppresses destructive updates to critical parameters to prevent overwriting old knowledge, while the latter explicitly characterizes feature shifts caused by condition changes and maintains a stable feature-space structure through self-calibration. Together, these mechanisms enhance model robustness and generalization under dynamic conditions. Experimental results demonstrate that DCEAN significantly outperforms representative methods in all phases, achieving smaller performance degradation, higher feature separability, and stronger stability in cross-phase tasks. Overall, DCEAN provides an effective and scalable modeling paradigm for exemplar-free incremental fault diagnosis under continuously variable operating conditions, offering important implications for intelligent health monitoring of industrial equipment under safety and privacy constraints. However, the current study does not yet address scenarios in which few-shot samples and varying operating conditions coincide with the emergence of previously unseen fault types, which constitutes an important direction for future research on DCEAN. Building on the demonstrated advantages of DCEAN in continual diagnosis under variable operating conditions, we are confident that, through further investigation and methodological refinement, it can be extended to effectively handle the diagnosis of unknown fault types in few-shot and variable-condition settings, thereby further enhancing the level of intelligent health monitoring for industrial equipment.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (No. 52272440), Suzhou Science Foundation (No. SYG202323), and Postgraduate Research and Practice Innovation Program of Jiangsu Province (KYCX25_3463).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

[1] B. Hou, Y. Wang, and D. Wang, "Investigations on multi-class classification model-based optimized weights spectrum for rotating machinery condition monitoring," *J. Dyn. Monit. Diagn.*, vol. 4, no. 3, pp. 194–202, 2025.

[2] Q. Li *et al.*, "Deep expert network: A unified method toward knowledge-informed fault diagnosis via fully interpretable neuro-symbolic AI," *J. Manuf. Syst.*, vol. 77, pp. 652–661, 2024.

[3] C. Shen *et al.*, "Pseudo-central feature matching: An adaptive semisupervised fault diagnosis method for knowledge transfer under variable working conditions," *Eng. Appl. Artif. Intell.*, vol. 163, p. 112875, 2026.

[4] K. Hu *et al.*, "An interpretable deep feature aggregation framework for machinery incremental fault diagnosis," *Adv. Eng. Inf.*, vol. 65, p. 103189, 2025.

[5] J. Xia *et al.*, "Digital twin-assisted fault diagnosis of rotating machinery without measured fault data," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–10, 2024.

[6] X. Wang *et al.*, "A trackable multi-domain collaborative generative adversarial network for rotating machinery fault diagnosis," *Mech. Syst. Sig. Process.*, vol. 224, p. 111950, 2025.

[7] Q. Li *et al.*, "Universal multimodal aggregation network with adaptive enhancement and semantic guidance for salient object detection," *Eng. Appl. Artif. Intell.*, vol. 156, p. 111371, 2025.

[8] Y. Chen *et al.*, "Class-consistent matching attention wavelet networks for partial transfer intelligent diagnosis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, pp. 15152–15166, 2025.

[9] A. Wan *et al.*, "A novel GA-PSO-SVM model for compound fault diagnosis in gearboxes with limited data," *IEEE Sens. J.*, vol. 25, pp. 30431–30443, 2025.

[10] R. Liu *et al.*, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Sig. Process.*, vol. 108, pp. 33–47, 2018.

[11] W. Sun, *et al.*, "LiteFormer: A lightweight and efficient transformer for rotating machine fault diagnosis," *IEEE Trans. Reliab.*, vol. 73, no. 2, pp. 1258–1269, 2023.

[12] Z. Lu *et al.*, "Rotating machinery fault diagnosis under multiple working conditions via a time-series transformer enhanced by convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.

[13] G. Yu *et al.*, "Few-shot fault diagnosis method of rotating machinery using novel MCGM based CNN, *IEEE Trans. Ind. Inf.*, vol. 19, no. 11, pp. 10944–10955, 2023.

[14] Z. Lei *et al.*, "SFUGDA: Source-free unsupervised multiscale graph domain adaptation network with privacy-preserving for cross-domain fault diagnosis of offshore wind turbines," *Mech. Syst. Sig. Process.*, vol. 235, p. 112896, 2025.

[15] J. Chen *et al.*, "An interpretable wavelet Kolmogorov–Arnold convolutional LSTM for spatial-temporal feature extraction and intelligent fault diagnosis," *J. Dyn. Monit. Diagn.*, vol. 4, no3, pp. 183–193, 2025.

[16] Y. Xiao *et al.*, "Domain generalization for rotating machinery fault diagnosis: A survey," *Adv. Eng. Inf.*, vol. 64, p. 103063, 2025.

[17] W. R. Sia *et al.*, "One-shot transfer learning with limited data sample for bearing component fault diagnosis," *Int. J. Hydromechatron.*, vol. 8, no. 6, pp. 1–29, 2025.

[18] Z. Wang *et al.*, "A generalized fault diagnosis framework for rotating machinery based on phase entropy," *Reliab. Eng. Syst. Saf.*, vol. 256, p. 110745, 2025.

[19] M. Niu, H. Jiang, and H. Shao, "Dynamic weighted adversarial domain adaptation network with sparse representation denoising module for rotating machinery fault diagnosis," *Eng. Appl. Artif. Intell.*, vol. 142, p. 109963, 2025.

[20] X. Wang *et al.*, "A dynamic collaborative adversarial domain adaptation network for unsupervised rotating machinery fault diagnosis," *Reliab. Eng. Syst. Saf.*, vol. 255, p. 110662, 2025.

[21] Y. Qin *et al.*, "Discriminative manifold domain adaptation for cross-domain fault diagnosis of rotating machineries," *Knowl.-Based Syst.*, vol. 285, p. 111332, 2024.

[22] X. Zhang *et al.*, "Adaptive variational sampling-embedded domain generalization network for fault diagnosis with intra-inter-domain class imbalance," *Reliab. Eng. Syst. Saf.*, 256, p. 110707, 2025.

[23] J. Wang *et al.*, "Multi-scale style generative and adversarial contrastive networks for single domain generalization fault diagnosis," *Reliab. Eng. Syst. Saf.*, vol. 243, p. 109879, 2024.

[24] Q. Song *et al.*, "Contrast-assisted domain-specificity-removal network for semi-supervised generalization fault diagnosis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 3, pp. 5403–5416, 2024.

[25] X. Li *et al.*, "Domain expansion fusion single-domain generalization framework for mechanical fault diagnosis under unknown working conditions," *Eng. Appl. Artif. Intell.*, vol. 138, p. 109380, 2024.

[26] X. Xu *et al.*, "A multi-sensor fused incremental broad learning with DS theory for online fault diagnosis of rotating machinery," *Adv. Eng. Inf.*, vol. 60, p. 102419, 2024.

[27] C. Shen *et al.*, "Dynamic branch layer fusion: A new continual learning method for rotating machinery fault diagnosis," *Knowl.-Based Syst.*, vol. 313, p. 113177, 2025.

[28] X. Tao *et al.*, "Semi-supervised feature contrast incremental learning framework for bearing fault diagnosis with limited labeled samples," *Appl. Soft Comput.*, vol. 176, p. 113172, 2025.

[29] J. Li *et al.*, "An uncertainty-aware continual learning framework for fault diagnosis of rotating machinery with homogeneous-heterogeneous faults, *IEEE Trans. Autom. Sci. Eng.*, pp. 1–15, 2024.

[30] H. Zhu *et al.*, "Reserving embedding space for new fault types: A new continual learning method for bearing fault diagnosis," *Reliab. Eng. Syst. Saf.*, vol. 252, p. 110433, 2024.

[31] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2935–2947, 2017.

[32] S.-A. Rebuffi *et al.*, "iCaRL: Incremental classifier and representation learning," *Proc. 2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 5533–5542, 2017.

[33] B. Zhao *et al.*, "Maintaining discrimination and fairness in class incremental learning," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit*., pp. 13208–13217, 2020.

[34] D. Goswami *et al.*, "Resurrecting old classes with new data for exemplar-free continual learning," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 28525–28534, 2024.