Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.077783

ARTICLE

Gearbox Fault Diagnosis under Varying Operating Conditions through Semi-Supervised Masked Contrastive Learning and Domain Adaptation

Zhixiang Huang[1,*] **and Jun Li**[1,2]

[1]School of Mechatronics and Vehicle Engineering, Chongqing Jiaotong University, Chongqing, China
[2]School of Electrical and Electronic Engineering, Chongqing Vocational and Technical University of Mechatronics, Chongqing, China

*Corresponding Author: Zhixiang Huang. Email: hzhixiang0513@163.com

Received: 16 December 2025; Accepted: 26 January 2026; Published: 26 February 2026

ABSTRACT: To address the issue of scarce labeled samples and operational condition variations that degrade the accuracy of fault diagnosis models in variable-condition gearbox fault diagnosis, this paper proposes a semi-supervised masked contrastive learning and domain adaptation (SSMCL-DA) method for gearbox fault diagnosis under variable conditions. Initially, during the unsupervised pre-training phase, a dual signal augmentation strategy is devised, which simultaneously applies random masking in the time domain and random scaling in the frequency domain to unlabeled samples, thereby constructing more challenging positive sample pairs to guide the encoder in learning intrinsic features robust to condition variations. Subsequently, a ConvNeXt-Transformer hybrid architecture is employed, integrating the superior local detail modeling capacity of ConvNeXt with the robust global perception capability of Transformer to enhance feature extraction in complex scenarios. Thereafter, a contrastive learning model is constructed with the optimization objective of maximizing feature similarity across different masked instances of the same sample, enabling the extraction of consistent features from multiple masked perspectives and reducing reliance on labeled data. In the final supervised fine-tuning phase, a multi-scale attention mechanism is incorporated for feature rectification, and a domain adaptation module combining Local Maximum Mean Discrepancy (LMMD) with adversarial learning is proposed. This module embodies a dual mechanism: LMMD facilitates fine-grained class-conditional alignment, compelling features of identical fault classes to converge across varying conditions, while the domain discriminator utilizes adversarial training to guide the feature extractor toward learning domain-invariant features. Working in concert, they markedly diminish feature distribution discrepancies induced by changes in load, rotational speed, and other factors, thereby boosting the model's adaptability to cross-condition scenarios. Experimental evaluations on the WT planetary gearbox dataset and the Case Western Reserve University (CWRU) bearing dataset demonstrate that the SSMCL-DA model effectively identifies multiple fault classes in gearboxes, with diagnostic performance substantially surpassing that of conventional methods. Under cross-condition scenarios, the model attains fault diagnosis accuracies of 99.21% for the WT planetary gearbox and 99.86% for the bearings, respectively. Furthermore, the model exhibits stable generalization capability in cross-device settings.

KEYWORDS: Gearbox; variable working conditions; fault diagnosis; semi-supervised masked contrastive learning; domain adaptation

1 Introduction

As a core component for transmitting mechanical power, gearboxes play a critical role in determining the operational efficiency and safety of mechanical systems [1]. In vital industries including new energy

vehicles, heavy machinery, and aerospace [2], gearboxes operating continuously under complex conditions are prone to performance degradation and eventual failure. Typical failure modes such as pitting, cracks, wear, and tooth breakage alter the system's dynamic response, potentially causing unscheduled downtime and serious safety incidents. Hence, achieving efficient and accurate gearbox fault diagnosis is of paramount practical importance.

The advancement and broad adoption of artificial intelligence have led to the extensive application of deep learning in mechanical fault diagnosis, owing to its powerful capabilities in automated feature extraction and pattern recognition. Representative deep learning architectures—including Convolutional Neural Networks (CNN) [3], Long Short-Term Memory networks (LSTM) [4], Transformer [5], and Gated Recurrent Units (GRU) [6]—hierarchically learn deep feature representations from vibration signals through multiple nonlinear transformations, demonstrating considerable strengths in gearbox fault diagnosis. For instance, Dutta et al. [7] introduced GearFaultNet, a lightweight 1D-CNN that effectively integrates multi-channel vibration data, attaining 94.04% accuracy in gearbox health state classification. Yang and Xu [8] proposed an approach integrating Extensive Empirical Wavelet Transform (EEWT), Entropy-Weighted Local Tangent Space Alignment (EWLTSA), and an Improved Deep Extreme Learning Machine (IDELM), reaching a diagnostic accuracy of 99.5%. He et al. [9] developed the MSCNN-LSTM-CBAM-SE model, which builds a more holistic fault representation through the deep fusion of multi-scale and temporal features.

However, these deep learning approaches generally depend on large volumes of high-quality, fully labeled fault samples for training and assume that training and test data follow identical distributions [10]. In reality, gearboxes operate in highly complex environments where conditions such as rotational speed and load fluctuate dynamically, resulting in considerable differences in the distribution of fault data across operating regimes—a phenomenon known as distribution shift [11]. Moreover, annotating high-quality samples demands substantial expert input, and samples of certain severe faults are inherently rare [12], making it increasingly difficult to construct large-scale, well-balanced labeled datasets. These issues collectively intensify the challenges of intelligent diagnosis under varying operating conditions, establishing limited-label fault diagnosis under such settings as a critical and persistent research challenge in both academic and industrial contexts.

In response to these challenges, researchers have pursued multiple directions. In feature extraction, Yang and Xu [8] introduced a hybrid method combining EEWT, EWLTSA, and IDELM, which extracts fault features robust to operational variations through adaptive signal decomposition and dimensionality reduction. Li et al. [13] built a semi-supervised framework based on adversarial training and pseudo-labeling, leveraging the dynamic interplay between generator and discriminator to uncover discriminative features from unlabeled data. In semi-supervised few-shot learning, Shao et al. [14] designed a revised loss function using label smoothing and metric scaling, improving cross-condition diagnosis under limited labels via refined similarity assessment. Yu et al. [15] integrated prototype networks with self-attention to extract discriminative fault features while capturing global dependencies. Lei et al. [16] devised a meta-transfer learning framework embedding prior knowledge as constraints during training. Tang et al. [17] combined coordinate attention with prototype calibration to achieve optimal cross-domain prototype alignment. Liang et al. [18] proposed a multi-scale multi-contrastive learning framework that sustains stable performance under extremely low label availability through hierarchical sample pairing. Nguyen et al. [12] developed the MLFork model, which enhances few-shot training, feature extraction, and pre-classification by capturing long-term dependencies via state-space modeling and localized vector attention, showing strong generalization under low-sample regimes, It is noteworthy that this design is based on a key premise: the target domain must possess a small amount of available true labeled data, with the core research question focusing on how to efficiently leverage these limited labels. Xu et al. [19] created the

FaultDiffusion framework, which pioneers the use of diffusion models to generate high-quality fault time-series samples via iterative denoising, offering effective data augmentation for few-shot learning. He et al. [20] innovatively leveraged pre-trained large language models to address the few-shot problem, proposing a GPT-2-based few-shot learning approach that fine-tunes the model by converting acoustic emission signal features into text, achieving high diagnostic accuracy for full ceramic bearings using only five labeled samples. In domain-conditioned feature correction, Li et al. [11] proposed a domain-adversarial graph convolutional network that minimizes structural discrepancies across domains using Maximum Mean Discrepancy. Wang et al. [21] designed a self-supervised pseudo-label learning method employing Cauchy MMD for global distribution alignment and an uncertainty-weighted confusion minimization strategy. Yang et al. [22] developed a cross-layer alignment network using joint MMD and norm constraints to strengthen domain adaptation. Cao et al. [23] built an unsupervised domain-shared CNN with an MMD variant, enabling knowledge transfer from steady to time-varying conditions. Zhou et al. [24] introduced an online transfer learning algorithm based on marginal probability distribution discrepancy, dynamically adapting to target domain shifts. Guo et al. [25] introduced a Deep Discriminative Adversarial Domain Adaptation Network (DDADAN), which effectively resolves the challenge of distribution discrepancy in cross-condition bearing fault diagnosis and enhances noise resistance by integrating wide-kernel convolutional feature extraction, correlation alignment-based adversarial training, and discriminative feature learning. Zhang and Gu [10] proposed a domain-conditioned feature correction approach that adjusts feature distributions to reduce operational condition-induced diagnostic errors.

However, current methods still exhibit the following shortcomings: ① Most approaches focus on the utilization of small samples, failing to fully exploit the latent value within vast amounts of unlabeled data; ② Existing semi-supervised contrastive learning methods rely on a substantial number of negative samples for contrastive learning, rendering model performance susceptible to the selection of these negative samples; ③ Furthermore, prevailing transfer learning techniques are incapable of precisely modeling inter-domain feature variability, which can easily lead models to concentrate on non-discriminative or redundant information, consequently degrading diagnostic accuracy and generalization capability. To address the aforementioned issues, this study proposes a gearbox fault diagnosis method based on SSMCL-DA. The principal innovations encompass: proposing a dual signal-augmented contrastive learning pre-training strategy to construct more challenging positive sample pairs; designing a ConvNeXt-Transformer hybrid feature extraction network as the backbone, which balances local feature extraction and long-range dependency capturing; and constructing a domain adaptation module that integrates Local Maximum Mean Discrepancy (LMMD) with adversarial learning to achieve fine-grained class-conditional distribution alignment, thereby furnishing a more comprehensive decision-making foundation for predictive maintenance in industrial settings. The effectiveness and generalization capability of the SSMCL-DA model are validated through multiple sets of experiments.

2 Basic Theory

The overall architecture of the SSMCL-DA model is illustrated in Fig. 1, and its workflow comprises two stages: pre-training and fine-tuning. During the pre-training stage, the model utilizes a ConvNeXt-Transformer encoder and adopts a dual signal-masked contrastive learning strategy for representation learning. In the fine-tuning stage, feature correction is first performed via a multi-scale attention mechanism. This mechanism serves as an implicit optimization module embedded within the forward propagation process, which simultaneously extracts macro-, meso-, and micro-scale physical features from the raw signal through parallel multi-scale convolutional kernels and subsequently applies temporal attention to dynamically weight the fused feature representations. Operating within a single forward pass, the mechanism

adaptively enhances critical fault segments, such as transient impulses, while suppressing noise and non-essential information, thereby significantly improving the model's feature discriminability and generalization performance under complex operating conditions. Following this, a domain adaptation module integrating Local Maximum Mean Discrepancy (LMMD) and adversarial learning is applied to reduce inter-domain distribution discrepancies.

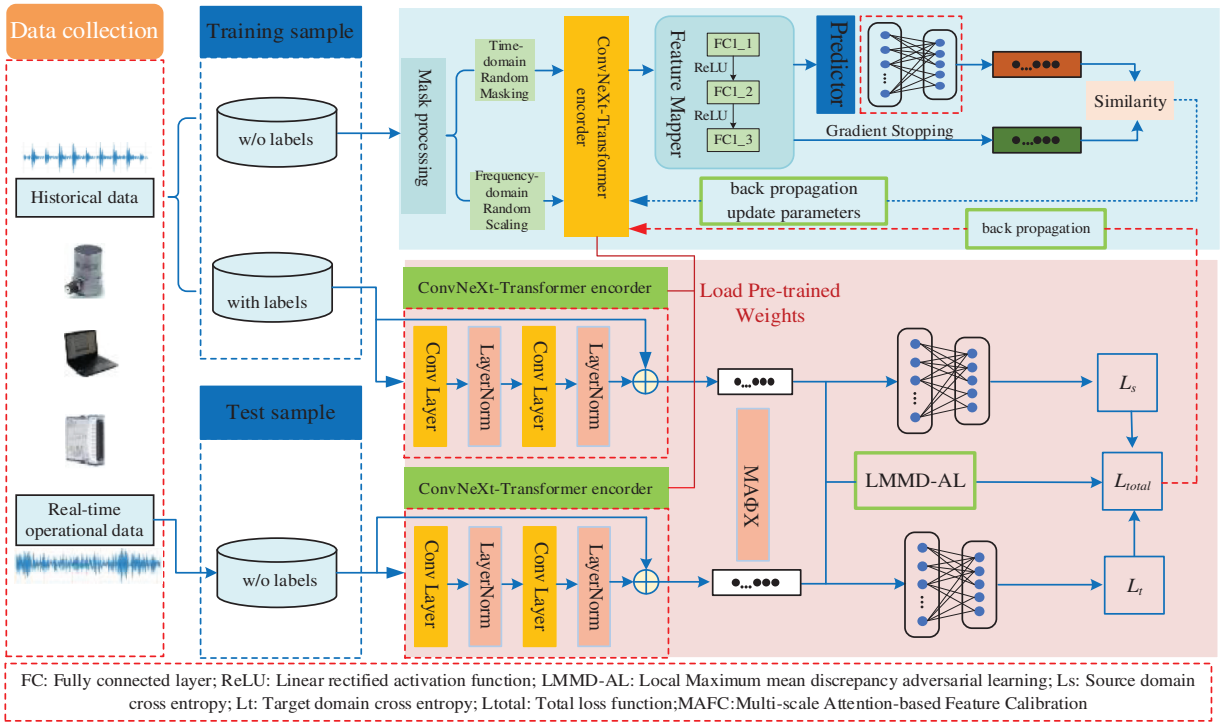

Figure 1: Network structure.

2.1 ConvNeXt-Transformer Encoder

Due to the absence of label guidance during pretraining on unlabeled data, the encoder must possess robust feature extraction capabilities to effectively capture and discriminate subtle differences in the data, thereby achieving superior feature representation learning in an unsupervised environment. Therefore, the SSM-CL model constructs a ConvNeXt-Transformer hybrid feature extraction network as the encoder, whose architecture is illustrated in Fig. 2. This network integrates the advantages of convolutional neural networks in local feature extraction with the strengths of Transformers in modeling global dependencies, enabling comprehensive capture of fault characteristics in vibration signals.

The lower layers of the network employ an improved ConvNeXt module to efficiently extract local time-domain/frequency-domain features. Through enhancements such as depthwise separable convolution, inverted bottleneck structures, and larger convolutional kernels, the ConvNeXt module significantly improves the model's representational capacity while maintaining the efficient local feature extraction capability of traditional CNNs. Its fundamental structure can be expressed as:

$$X_l + 1 = ConvNeXt\left(LN\left(X_l\right)\right) + X_l \tag{1}$$

here, X_l denotes the input features to the l-th layer, LN represents layer normalization, and the *ConvNeXt* block comprises a composite operation of depthwise convolution and pointwise convolution.

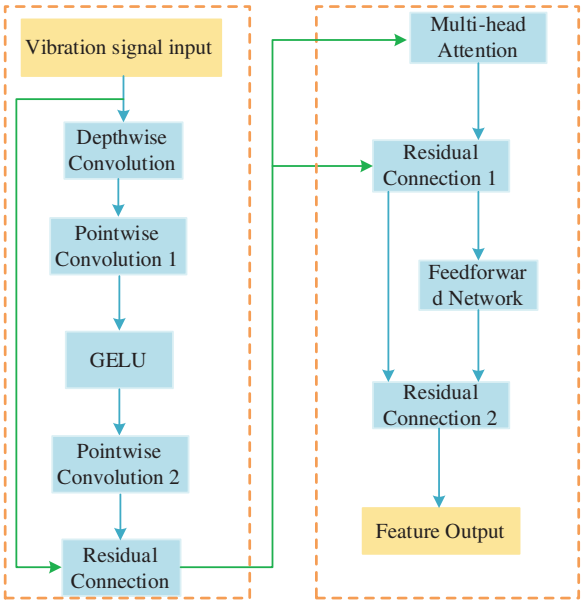

Figure 2: Schematic architecture of ConvNeXt-transformer hybrid feature extraction network.

The higher network layers incorporate Transformer encoder layers, leveraging self-attention mechanisms to capture long-range dependencies among fault-induced impulses in vibration signals. By establishing connections across all input positions, the self-attention mechanism effectively models global contextual information within the signal. The computational procedure is as follows:

$$Attention\,(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{dk}}\right)V \tag{2}$$

here, Q, K and V denote the query, key, and value matrices, respectively, and d_k represents the dimensionality of the key vectors. Through the multi-head self-attention mechanism, the model can jointly attend to information from different positions across multiple representation subspaces.

This hybrid architecture fully leverages the efficiency of CNNs in local feature extraction and the strength of Transformers in modeling global dependencies, forming a complementary feature learning mechanism. The ConvNeXt module is responsible for extracting discriminative local patterns from raw vibration signals, while the Transformer layer integrates these local features and establishes global contextual relationships. This synergy enhances feature representation capability for complex vibration signals and establishes a solid foundation for subsequent domain adaptation and classification tasks.

2.2 Dual-Signal-Augmented Masked Contrastive Learning

Dual-signal-augmented masked contrastive learning represents a self-supervised methodology whose core concept involves simultaneously applying both temporal masking and frequency-domain perturbation augmentation strategies to input data, while integrating a contrastive learning mechanism to construct optimization objectives within unlabeled datasets, thereby prompting the model to learn more discriminative and robust feature representations.

Dual-Signal Augmentation Strategy. For each input sample $x \in \mathbb{R}^L$ (where L denotes the sequence length), two independent signal augmentation operations are simultaneously applied:

2.2.1 Temporal Masking Augmentation

A random masking operation is employed, setting randomly selected data points within the input sequence to zero, thereby compelling the model to learn contextual information for reconstructing the masked content. The masked data is expressed as:

$$x_{mask} = x \odot M \tag{3}$$

where $M \in \{0,1\}^L$ is a masking matrix whose elements satisfy:

$$M_i = \begin{cases} 0 \; with \; probability & p_{mask} \\ 1 \; otherwise \end{cases} \tag{4}$$

here p_{mask} denotes the masking probability, controlling the masking ratio.

2.2.2 Frequency-Domain Perturbation Augmentation

The input signal undergoes Fast Fourier Transform (FFT), with its magnitude spectrum subjected to random scaling and phase perturbation, simulating minor rotational speed variations and phase jitter encountered in practical applications:

Perform FFT on the original signal: $X = FFT(x)$.

Compute the magnitude spectrum $A = |X|$ and phase spectrum $\phi = \angle X$.

Apply random perturbations to magnitude and phase:

$$\tilde{A} = \alpha \cdot A, \quad \tilde{\phi} = \phi + \delta \tag{5}$$

where $\alpha \sim u(0.8, 1.2), \quad \delta \sim u(-0.1\pi, 0.1\pi)$.

Reconstruct the signal using perturbed spectra:

$$x_{freq} = IFFT\left(\tilde{A} \cdot e^{j\tilde{\phi}}\right) \tag{6}$$

2.2.3 Contrastive Learning Framework

Through the aforementioned augmentations, we obtain a pair of positive sample views for each sample x:

$$v_1 = x_{mask}, v_2 = x_{freq} \tag{7}$$

They are fed into a shared-weight encoder f_θ to extract features:

$$z_1 = f_\theta(v1), z_2 = f_\theta(v_2) \tag{8}$$

Subsequently, feature vectors z_1, z_2 are mapped via a projection head g_ϕ (typically a small MLP) to a space more suitable for contrastive learning:

$$h_1 = g_\phi(z_1), h_2 = g_\phi(z_2) \tag{9}$$

To enable asymmetric prediction and avoid trivial solutions, a prediction head q_ψ (another MLP) is introduced, computing:

$$p_1 = q_\psi(h_1), p_2 = q_\psi(h_2) \tag{10}$$

A symmetric contrastive loss function is adopted, whose core objective is to maximize the similarity between the prediction vectors of two augmented views and each other's projection vectors. The negative cosine similarity is defined as the distance metric:

$$D(u, v) = -\frac{u \times v}{\| v \|_2 \| v \|_2} \tag{11}$$

To prevent model collapse, gradient stopping operation $sg(\cdot)$ is applied to the target vectors when computing the loss. The final symmetric loss function is:

$$\mathcal{L} = \frac{1}{2}\left[\mathcal{D}(p_1, sg(h_2)) + \mathcal{D}(p_2, sg(h_1))\right] sg(\cdot) \tag{12}$$

By jointly employing temporal masking and frequency-domain augmentation, positive sample pairs that closely resemble actual operating conditions (such as rotational speed fluctuations) are constructed. This dual challenge forces the encoder to abandon reliance on single-dimensional features, thereby learning highly robust intrinsic features that are insensitive to both temporal incompleteness and frequency variations.

2.3 Domain-Conditioned Feature Correction with Local Maximum Mean Discrepancy (LMMD) and Adversarial Learning

The domain-conditioned feature correction strategy primarily reduces inter-domain distribution discrepancies by identifying domain-invariant features [11]. Its objective is to leverage labeled source domain data and unlabeled target domain data during the training phase, adopting a joint alignment strategy that combines Local Maximum Mean Discrepancy (LMMD) with adversarial learning, thereby compelling the model to learn condition-invariant features. This ensures that feature distributions for identical faults under different operating conditions (such as load and rotational speed variations) converge toward consistency, enhancing the model's cross-domain generalization capability.

For the source domain features $H_l(x_s)$ and target domain features $H_l(x_s)$ output from the l-th $(l = 1, 2, \ldots, L)$ layer of the model, a domain-conditioned feature correction strategy is introduced to adjust the target domain features. The domain-conditioned feature correction module comprises a fully connected layer, a ReLU activation function, and another fully connected layer, generating feature correction terms for the target domain. The corrected target domain features are expressed as:

$$\hat{H}_l(x_t) = H_l(x_t) + \Delta H_l(x_t) \tag{13}$$

where $\Delta H_l(x_t)$ represents the feature correction term for the target domain.

To more comprehensively reduce inter-domain discrepancies, we introduce a domain adaptation module combining LMMD (Local Maximum Mean Discrepancy) with adversarial learning. LMMD is responsible for aligning feature distributions of the same fault category between source and target domains (inter-class alignment), with its empirical estimation formula given by:

$$L_{LMMD} = \frac{1}{C}\sum_{c=1}^{C}\left\| \frac{1}{n_s^c}\sum_{i=1}^{n_s^c}\phi\left(H_l\left(x_s^i\right)\right) - \frac{1}{n_t^c}\sum_{j=1}^{n_t^c}\phi\left(H_l\left(x_t^i\right)\right)\right\|_{\mathcal{H}_k}^2 \tag{14}$$

where C is the number of fault categories, n_s^c and n_t^c denote the sample counts belonging to category c in the source and target domains, respectively, $\mathcal{H}_k$ represents the reproducing kernel Hilbert space with characteristic kernel k, and ϕ denotes the feature mapping.

Simultaneously, a domain discriminator D is introduced for adversarial training with the feature extractor. The domain discriminator attempts to distinguish whether features originate from the source or target domain, while the feature extractor aims to "deceive" the discriminator, thereby learning global domain-invariant features (global alignment). The loss function for the domain discriminator is:

$$L_{adv}^{D} = -\mathbb{E}_{x_s \sim \mathcal{S}} \left[\log D \left(H_l \left(x_s \right) \right) \right] - \mathbb{E}_{x_t \sim \mathcal{T}} \left[\log \left(1 - D \left(H_l \left(x_t \right) \right) \right) \right] \tag{15}$$

The adversarial loss for the feature extractor is:

$$L_{adv}^{F} = -\mathbb{E}_{x_t \sim \mathcal{T}} \left[\log D \left(H_l \left(x_t \right) \right) \right] \tag{16}$$

Simultaneously, to prevent excessive migration by the domain-conditioned feature correction block, feature correction is also applied to source data for regularization. Typically, the source domain feature representation should remain unchanged after passing through the feature correction block. However, perfect alignment of each category in the source domain would lead to $\Delta H_l \left(x_s \right) \approx 0$, indicating that the correction block learns nothing in cross-domain feature correction learning. Therefore, a regularization loss is introduced to address this issue. For a random subset of source data, used to appropriately guide the correction process and enhance the alignment capability of the correction block, the regularization loss is defined as follows:

$$L_{reg} = \frac{1}{|R|} \sum_{x_s \in R} \left\| \Delta H_l \left(x_s \right) \right\|_2^2 \tag{17}$$

where R represents a random subset of the source domain, $|R|$ denotes the subset size, and the probability for each data point being randomly selected is defined as P/C_n, with P being a control factor.

In unsupervised domain adaptation, the objective is to learn a classifier with strong generalization capability on the target domain by exploring labeled source data and unlabeled target data during the training phase. Since only the source data possesses labels, a source classifier can be constructed by minimizing the following loss function:

$$L_s = \frac{1}{n_s} \sum_{i=1}^{n_s} \mathcal{L}_{ce} \left(G \left(H \left(x_s^i \right) \right), y_s^i \right) \tag{18}$$

where $\mathcal{L}_{ce} \left(\cdot, \cdot \right)$ denotes the cross-entropy loss function and $G \left(\cdot \right)$ represents the learned predictive model. However, since this loss only learns a representation mapping sensitive to the source domain while domain distribution discrepancies exist, the source classifier may not generalize effectively to the target domain. Moreover, as the target domain is unlabeled, it is reasonable to adopt the entropy minimization principle to enhance the discriminative power of the learned model. Defining the conditional probability of the k-th class for target domain data x_t predicted by $G \left(\cdot \right)$ as $G_k \left(x_t \right)$, the target domain entropy loss function is then:

$$L_e = -\frac{1}{n_t} \sum_{j=1}^{n_t} \sum_{k=1}^{C} G_k \left(x_t^j \right) \log G_k \left(x_t^j \right) \tag{19}$$

Therefore, the overall optimization objective for domain-conditioned correction can be expressed as:

$$\min_{G,F} \max_{D} L = L_s + \lambda L_e + \alpha \sum_{l=1}^{L} L_{LMMD}^{l} + \beta \sum_{l=1}^{L} L_{adv}^{F,l} + \gamma \sum_{l=1}^{L} L_{reg}^{l} \tag{20}$$

where L_s and L_e represent the source domain classification loss and target domain entropy loss, respectively, L^l_{LMMD} and $L^{F,l}_{adv}$ denote the *LMMD* alignment loss and adversarial loss at the l-th layer, respectively, L^l_{reg} indicates the regularization loss, and α, β, γ and λ are weighting factors. This joint alignment mechanism combining *LMMD* with adversarial learning, through the complementarity of fine-grained class-center alignment and coarse-grained global distribution alignment, can more comprehensively and stably reduce inter-domain discrepancies and enhance transfer performance.

It is important to note that, in the practical training process, we adopt an integrated implementation strategy to enhance optimization efficiency and model robustness: the classification loss weight ($\alpha = 1$) is fixed as the optimization baseline; the *LMMD* loss (β) and the adversarial loss (γ) are unified into a single domain adaptation objective, whose overall intensity is dynamically managed by an adaptive curriculum learning strategy, enabling the model to autonomously balance the focus between classification and alignment; the constraining effect of the regularization term (λ) is primarily ensured by the model structure itself (e.g., through normalization layers). This design ensures the model's powerful diagnostic performance while significantly reducing the number of hyperparameters requiring manual tuning, thereby enhancing the method's practicality and reproducibility.

3 Gearbox Fault Diagnosis under Variable Operating Conditions Based on Semi-Supervised Masked Contrastive Learning and Domain Adaptation (SSMCL-DA)

The SSMCL-DA process framework is presented in Fig. 3.

(1) For a substantial volume of unlabeled vibration samples, a dual signal enhancement strategy—comprising random masking in the time domain and random scaling in the frequency domain—is employed to generate two augmented views for each sample. These two views are subsequently input into a ConvNeXt-Transformer hybrid encoder to extract their corresponding feature representations.

(2) Based on the extracted features, a contrastive loss is constructed with the objective of maximizing feature similarity between different augmented views derived from the same original sample. The encoder parameters are updated via backpropagation, enabling the model to learn robust and consistent intrinsic features from unlabeled data that are invariant to variations in operating conditions.

(3) Upon reaching the predetermined number of training epochs, the encoder weights are saved, thereby completing the unsupervised pre-training stage.

(4) The model is fine-tuned utilizing a limited amount of labeled source domain data and unlabeled target domain data. Feature calibration is performed through a multi-scale attention mechanism, while joint optimization of the Local Maximum Mean Discrepancy (LMMD) loss and the adversarial domain discrimination loss is conducted to achieve fine-grained class-conditional distribution alignment and global domain-invariant feature learning.

(5) The target domain data to be diagnosed is input into the trained model, and the classifier outputs the final fault category prediction, thereby accomplishing cross-condition fault diagnosis.

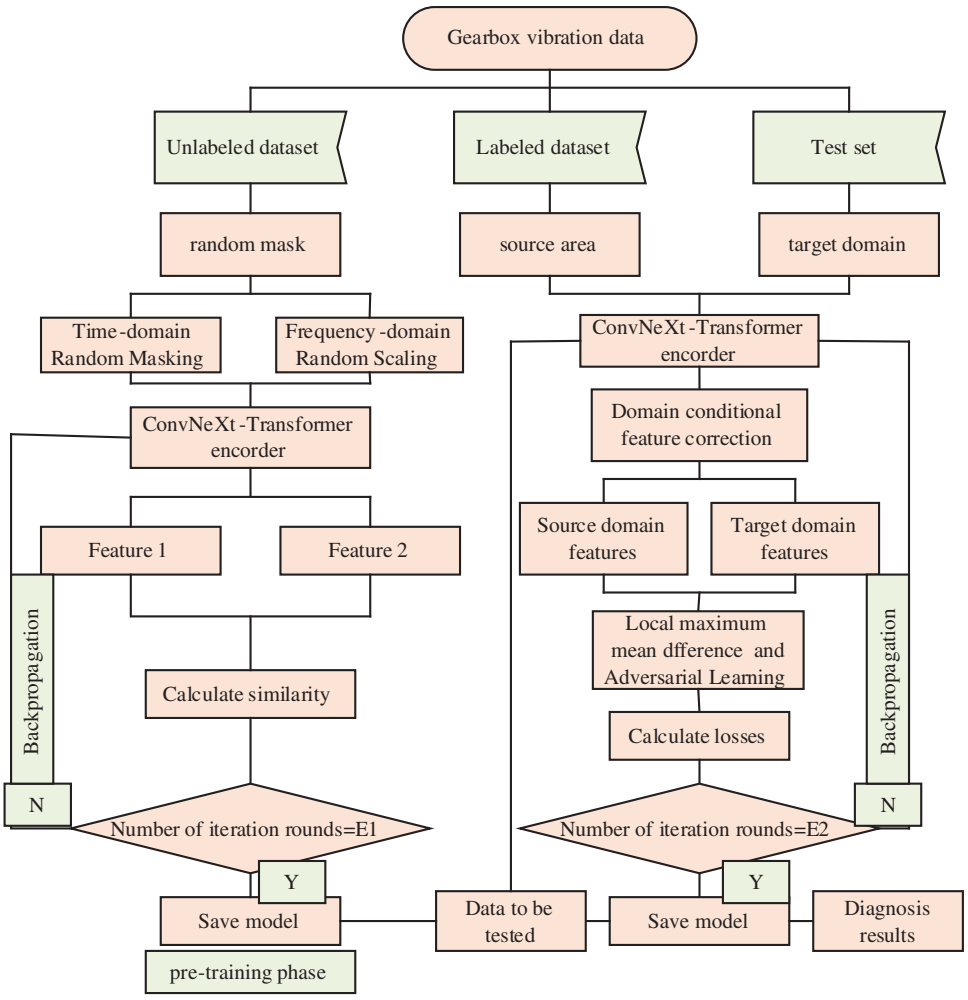

Figure 3: Framework diagram of the SSMCL-DA model fault diagnosis process.

4 Experimental Analysis

4.1 Experimental Data

The WT-Planetary Gearbox dataset [26] employed in this study is collected from a wind turbine transmission system experimental testbed (Fig. 4). The testbed consists of a motor, a planetary gearbox, a fixed-shaft gearbox, and a loading device, wherein the planetary gearbox is composed of a sun gear, four planet gears, and a ring gear. The dataset is collected for five health conditions of the sun gear (healthy, tooth breakage, wear, root crack, and missing tooth). The vibration signals are measured simultaneously by two orthogonally oriented Sinocera CA-YD-1181 accelerometers, with the rotational speed pulses of the input shaft being simultaneously recorded via an encoder at a sampling frequency of 48 kHz. The continuous acquisition time for each health condition exceeds 5 min, ensuring sufficient samples are generated to meet the training requirements of deep learning models. The dataset includes data under eight different input shaft rotational speeds (20–55 Hz), making it suitable for research on cross-condition generalization and transfer learning of models. Furthermore, all data were repeatedly collected after complete disassembly and reassembly of the planetary gearbox, providing a unique basis for evaluating the robustness of models to "installation effects".

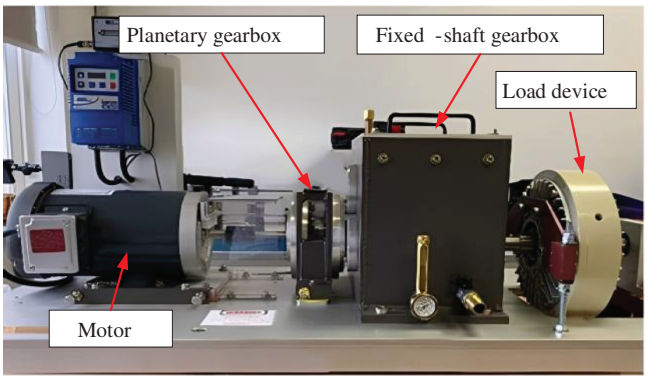

Figure 4: Wind turbine drive train test rig.

The operational conditions of the gearbox are configured as 20 Hz (Condition A), 25 Hz (Condition B), and 30 Hz (Condition C). To validate the effectiveness and advantages of the proposed method, fault diagnosis experiments with varying labeled sample sizes and small-sample fault diagnosis experiments under identical conditions were conducted. The partitioning of the experimental dataset is detailed in Table 1, where each sample consists of 1024 data points. The dataset is divided into training samples and test samples, with the training set further comprising unlabeled data and a limited amount of labeled data for the pre-training and fine-tuning stages of model training.

Table 1: Gearbox dataset.

Labels	Fault type	Training set		Test set
		Without labels	With labels	With outlabels
0	Normal	400	80	120
1	Pitting	400	80	120
2	Tooth breakage	400	80	120
3	Crack	400	80	120
4	Wear	400	80	120
Total		2000	400	600

4.2 Parameterization

The SSMCL-DA model employs the Adam optimizer during the pre-training phase, with the learning rate set to 0.0001 and a fixed learning rate schedule, training for 2 epochs with a batch size of 64. In the fine-tuning stage, the Adam optimizer is also used, with a fixed learning rate of 0.0001, training for 50 epochs and a batch size of 64. The detailed parameters of the ConvNeXt-Transformer feature extraction encoder are listed in Table 2. The experiments were conducted on the deep learning framework PyTorch 1.11, with a hardware environment consisting of an Intel Core i7-11800H CPU and an NVIDIA RTX3070 GPU. To demonstrate the superiority of the SSMCL-DA model, comparative experiments were performed against CNN-LSTM, CNN-BiGRU, CNN-Transformer, and MSCNN-LSTM-Attention models under different diagnostic task scenarios. The performance of each model was holistically evaluated based on diagnostic accuracy and feature visualization plots.

Table 2: Model parameter.

Model	Operation	Size/Step size
Patch Embedding	Conv2d	Kernel = 4 × 4, Stride = 4, Channels = 96
Stage 1	ConvNeXt Block × 3	Input = 96, Output = 96
Downsampling	LayerNorm + Conv2d	Kernel = 2 × 2, Stride = 2, Output = 192
Stage 2	ConvNeXt Block × 3	Input = 192, Output = 192
Downsampling	LayerNorm + Conv2d	Kernel = 2 × 2, Stride = 2, Output = 384
Stage 3	ConvNeXt Block × 9	Input = 384, Output = 384
Downsampling	LayerNorm + Conv2d	Kernel = 2 × 2, Stride = 2, Output = 768
Stage 4	ConvNeXt Block × 3	Input = 768, Output = 768

4.3 Fault Diagnosis Experiments under the Same Working Condition with Different Label Sample Sizes

To validate the effectiveness of the SSMCL-DA model in gearbox fault diagnosis, a comparative experiment was conducted between the proposed method and several benchmark methods under identical operating conditions but with varying quantities of labeled samples. The experimental setup comprised three distinct operating conditions, with five different labeled sample quantities tested under each condition: 20, 40, 60, 80, and 100. Table 3 presents the classification accuracies of all methods under the same operating condition, while Fig. 5 illustrates the accuracy progression curves of each method during the iterative process when the labeled sample size is 80 (e.g., for the A → A operating condition).

Table 3: Classification accuracy at the same speed (%).

Working condition	Methods	Number of label samples				
		20	**40**	**60**	**80**	**100**
A → A	CNN-LSTM	93.48	94.08	94.78	95.08	95.91
	CNN-BiGRU	91.43	92.35	92.87	93.65	95.01
	CNN-Transformer	77.39	80.18	81.35	82.99	83.07
	MSCNN-LSTM-Attention	97.28	98.02	98.47	98.97	99.01
	SSMCL-DA	98.34	98.76	99.24	99.80	99.87
B → B	CNN-LSTM	91.32	91.88	93.22	94.18	96.03
	CNN-BiGRU	88.24	91.28	94.37	94.88	96.31
	CNN-Transformer	75.46	77.39	80.55	82.17	87.23
	MSCNN-LSTM-Attention	95.33	97.36	97.47	97.93	98.11
	SSMCL-DA	96.83	97.96	99.24	99.33	99.57
C → C	CNN-LSTM	85.34	88.48	90.23	92.01	93.11
	CNN-BiGRU	83.76	88.32	90.37	92.05	92.81
	CNN-Transformer	73.56	74.13	75.31	77.36	79.98
	MSCNN-LSTM-Attention	97.28	98.02	98.47	98.97	99.01
	SSMCL-DA	97.44	98.03	98.81	99.04	99.18

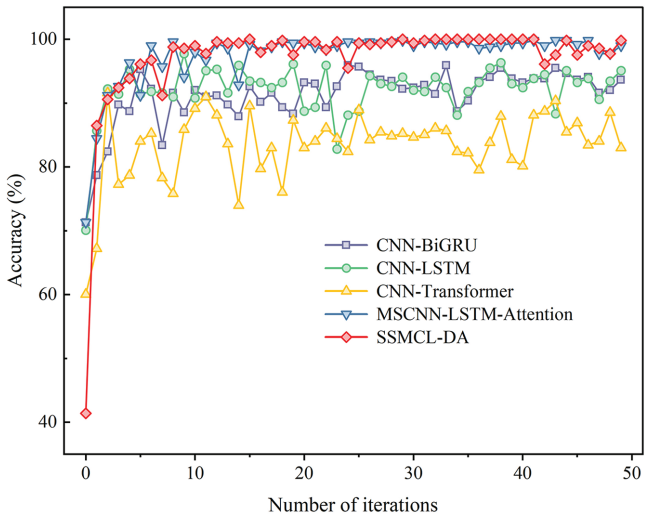

Figure 5: A → A comparison of the results of each method with a label sample size of 80.

As indicated in Table 3 and Fig. 5, as the number of labeled samples available for training decreases, the diagnostic accuracy of all models declines to varying degrees. Taking the A → A operating condition as an example, when the labeled sample size is 80, although the accuracy of all models remains above 80%, allowing for fundamentally effective classification, significant performance discrepancies exist among the models. Among them, the CNN-Transformer model exhibits the lowest diagnostic accuracy (82.99%), and its iterative accuracy curve shows pronounced fluctuations, hindering stable convergence. This is primarily attributed to the model's tendency to overfit specific patterns within the training samples under label scarcity, failing to adequately learn the overall distribution characteristics of the data. In contrast, the MSCNN-LSTM-Attention and CNN-LSTM models achieve superior classification performance by leveraging their enhanced sequential modeling and feature fusion capabilities, improving accuracy by 16.81% and 12.09%, respectively, compared to the CNN-Transformer model. Notably, the MSCNN-LSTM-Attention model, through the synergistic mechanism of multi-scale feature extraction, deep sequence comprehension, and dynamic attention-weight focusing, achieves more precise and robust modeling of complex spatiotemporal patterns, attaining an accuracy of 98.87%, which approaches that of the proposed SSMCL-DA model (99.80%). This result demonstrates that, under identical operating conditions, both MSCNN-LSTM-Attention and SSMCL-DA exhibit strong feature extraction capabilities. However, SSMCL-DA maintains a slight yet consistent advantage, fundamentally due to its adoption of dual-signal-augmented masked contrastive learning and a ConvNeXt-Transformer hybrid architecture. During the pre-training phase, the application of time-domain masking and frequency-domain scaling to unlabeled samples constructs more challenging positive sample pairs, driving the encoder to learn intrinsic features robust to operational variations while fully exploiting the latent information embedded in the unlabeled data. Concurrently, the hybrid architecture effectively fuses the local detail modeling prowess of ConvNeXt with the global perceptual capacity of Transformer, thereby achieving superior feature representation and extraction efficacy.

Additionally, to further validate the feature extraction capability of the SSMCL-DA model, the t-distributed Stochastic Neighbor Embedding (t-SNE) method was employed to reduce the dimensionality and visualize the features output from the fully connected layers of each method under the A → A condition with 80 labeled samples; the results are shown in Fig. 6. As illustrated in Fig. 6a, the raw data features exhibit a chaotic distribution where features from different categories overlap and remain unseparated. As shown in Fig. 6b–f, after feature extraction by each respective method, the dimensionality-reduced features all

display certain classification boundaries; however, due to the limited sample size, varying degrees of inter-category overlap persist across all methods. Specifically, a small number of sample points are misclassified in the MSCNN-LSTM-Attention model, while the SSMCL-DA model exhibits only minor confusion between individual points (such as between Broken Tooth and Missing Tooth), with clear separation boundaries evident among all other categories. This further confirms that the SSMCL-DA model maintains excellent feature discrimination and extraction capability even in small-sample scenarios.

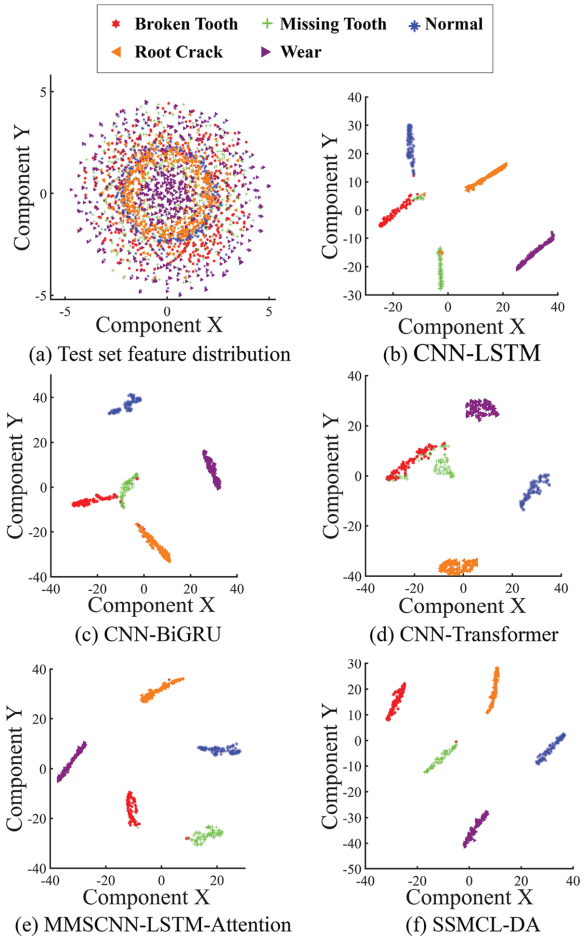

Figure 6: A → A comparison of the feature distribution of each method with a label sample size of 80.

4.4 Fault Diagnosis Experiments under Variable Operating Conditions with Different Label Sample Sizes

To further validate the effectiveness of the SSMCL-DA model for fault diagnosis under variable operating conditions with different labeled sample sizes, SSMCL-DA was compared with benchmark algorithms through experiments involving six distinct transfer tasks and five varying labeled sample quantities. All experiments were conducted on a unified hardware platform comprising an Intel Core i7-11800H CPU and an NVIDIA RTX 3070 GPU. The results demonstrate that completing a typical cross-condition transfer task, such as B → C, for a total of 52 epochs requires only approximately 4 to 5 min, which highlights the framework's efficient training iteration capability. Simultaneously, the final streamlined inference model deployed for practical application—which retains only the fixed ConvNeXt-Transformer backbone network and the classifier—exhibits excellent lightweight characteristics: its parameter count is approximately 21.5M, its weight file size is merely 6.06 MB, and its average inference latency for a single sample (1024 data points)

remains below 5 ms, thereby establishing an efficiency foundation for the engineering implementation of the core algorithm.

Table 4 presents the diagnostic accuracy under variable operating conditions, while Fig. 7 depicts the iterative accuracy curves of each method for the transfer task B → C with a labeled sample size of 80. As shown in Table 4, under variable operating conditions, all models except SSMCL-DA exhibit a significant performance degradation compared to their performance under identical conditions. Specifically, the accuracy of the CNN-LSTM and CNN-Transformer models decreases by 16.19% and 10.04%, respectively. The MSCNN-LSTM-Attention model, due to its failure to accurately model and align inter-domain feature discrepancies, causes the model to focus on non-discriminative or redundant information, resulting in a diagnostic accuracy (94.38%) slightly lower than that of the SSMCL-DA model (99.18%). This is primarily attributable to data distribution shifts caused by operational differences and the domain adaptation challenges posed by insufficient labeled samples. In contrast, the SSMCL-DA model employs a multi-scale attention mechanism for feature correction, achieving fine-grained distribution alignment under class-specific conditions and learning domain-invariant features. Consequently, it attains the highest average diagnostic accuracy across all transfer tasks. Furthermore, as illustrated in Fig. 7, in the B → C task with a labeled sample size of 80, SSMCL-DA demonstrates the fastest convergence, stabilizing around the 8th iteration, which further validates the effectiveness and robustness of the proposed method for fault diagnosis tasks under variable operating conditions.

Table 4: Accuracy under variable operating conditions (%).

Number of label samples	Methods	Different speed						
		A → B	A → C	B → C	B → A	C → A	C → B	Average
20	CNN-LSTM	70.81	71.28	75.25	75.17	71.08	72.13	72.62
	CNN-BiGRU	76.33	79.41	86.03	86.38	85.79	84.22	83.03
	CNN-Transformer	68.39	68.00	70.13	69.11	72.88	71.35	69.98
	MSCNN-LSTM-Attention	89.77	85.92	91.22	91.34	89.21	90.37	89.64
	SSMCL-DA	97.28	97.73	97.06	98.03	96.89	97.21	97.37
40	CNN-LSTM	71.10	71.88	76.33	76.11	72.14	73.24	73.47
	CNN-BiGRU	76.87	80.02	86.34	86.93	86.09	85.08	83.56
	CNN-Transformer	69.89	69.04	70.56	69.88	73.19	73.21	70.96
	MSCNN-LSTM-Attention	90.18	86.08	91.89	92.18	90.11	91.37	90.30
	SSMCL-DA	97.39	97.28	97.89	99.14	96.96	97.75	97.74
60	CNN-LSTM	72.89	72.43	78.08	77.46	74.28	74.58	74.95
	CNN-BiGRU	78.04	81.11	86.55	87.28	87.35	87.35	84.61
	CNN-Transformer	77.17	70.19	71.64	70.22	74.56	74.59	73.06
	MSCNN-LSTM-Attention	90.12	87.17	92.47	93.46	91.27	92.33	91.14
	SSMCL-DA	97.55	97.78	98.32	99.15	97.24	98.31	98.06
80	CNN-LSTM	73.11	73.84	78.89	78.09	75.66	75.19	75.80
	CNN-BiGRU	79.01	82.67	87.09	88.33	88.39	88.15	85.61
	CNN-Transformer	78.23	71.49	72.95	71.28	76.22	75.63	74.30

(Continued)

Table 4 (continued)

Number of label samples	Methods	Different speed						
		A → B	A → C	B → C	B → A	C → A	C → B	Average
	MSCNN-LSTM-Attention	91.92	88.27	94.38	94.17	92.38	94.28	92.57
	SSMCL-DA	97.81	97.87	99.18	99.18	97.94	98.50	98.41
100	CNN-LSTM	74.37	74.08	80.12	79.34	77.68	77.14	77.12
	CNN-BiGRU	67.75	83.47	88.05	89.37	89.61	89.33	84.60
	CNN-Transformer	79.25	72.14	73.21	72.64	77.65	76.88	75.30
	MSCNN-LSTM-Attention	92.14	89.16	95.28	95.37	93.12	95.04	93.35
	SSMCL-DA	97.95	98.44	99.21	99.20	97.84	99.14	98.63

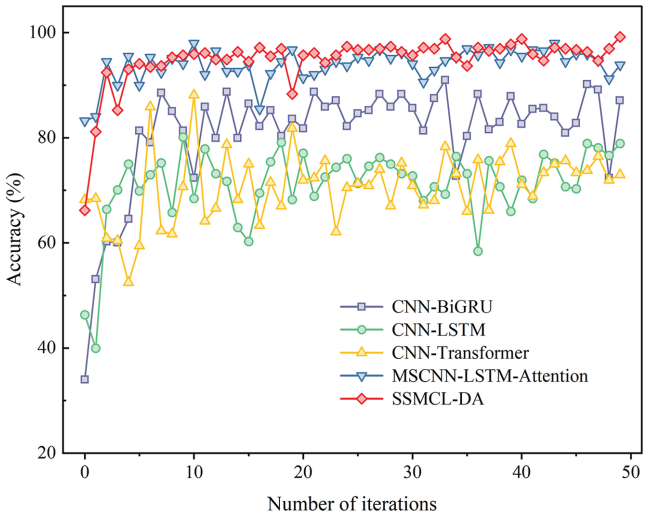

Figure 7: B → C comparison of the results of each method with a label sample size of 80.

To validate that the SSMCL-DA model can effectively reduce the distribution discrepancy between the source and target domains, this study takes the B → C transfer task (with a labeled sample size of 80) as an example and comparatively analyzes the Probability Density Plots (PDF) and Cumulative Distribution Functions (CDF) of the MSCNN-LSTM-Attention and SSMCL-DA models, with the results presented in Fig. 8.

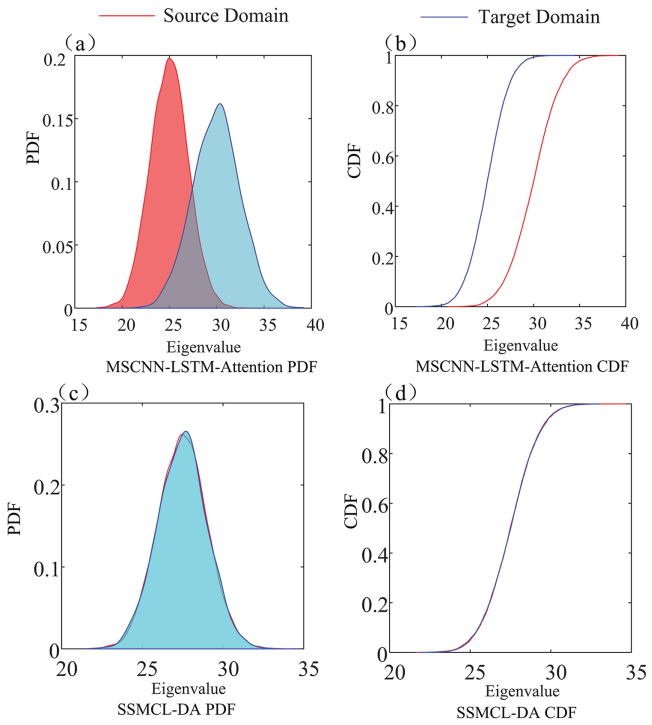

Figure 8: Probability density (PDF) and cumulative distribution (CDF) plots for the two models: (**a**) MSCNN-LSTM-Attention PDF, (**b**) MSCNN-LSTM-Attention CDF, (**c**) SSMCL-DA PDF, and (**d**) SSMCL-DA CDF.

As observed in Fig. 8a,b, the MSCNN-LSTM-Attention model exhibits a significant discrepancy between the data distributions of the source and target domains, where neither the probability density curves nor the CDFs are aligned. This visually reflects the inter-domain distribution shift induced by variations in operating conditions. In contrast, the SSMCL-DA model, through feature calibration via its multi-scale attention mechanism and the joint domain adaptation module that integrates Local Maximum Mean Discrepancy (LMMD) with adversarial learning, significantly narrows the distribution gap between the target and source domains. As shown in Fig. 8c,d, the Probability Density Plots (PDF) and Cumulative Distribution Functions (CDF) of the source and target domains in the SSMCL-DA model demonstrate a high degree of overlap, which verifies the effectiveness of the model in cross-domain feature alignment and highlights the superior performance of its domain adaptation mechanism.

To verify the effectiveness of each core component in the SSMCL-DA model, a systematic ablation study was conducted under the transfer task B → C (with a labeled sample size of 80). The experimental results, presented in Table 5, demonstrate the following: At the feature extraction level, merely replacing the CNN encoder with the ConvNeXt-Transformer encoder enhanced the accuracy by 9.53%, demonstrating its powerful representational capacity. At the algorithmic level, the individual introduction of dual-signal masked contrastive learning or the LMMD+ adversarial domain adaptation module yielded performance improvements of 5.54% and 9.98%, respectively, with the domain adaptation method showing more pronounced effectiveness. This clearly indicates that mitigating inter-domain distribution discrepancies is more critical than solely enhancing feature discriminability in this cross-domain diagnostic task. Regarding module synergy, integrating both algorithms with the CNN encoder led to an accuracy improvement of 13.86%; whereas combining ConvNeXt-Transformer with the core domain adaptation module further increased the gain to 14.79%. The complete model ultimately achieved the optimal accuracy of 99.18%,

representing an overall improvement of 18.83%. This fully validates the complementary nature of the modules: ConvNeXt-Transformer enhances feature representation capability, masked contrastive learning promotes the extraction of discriminative features, and adversarial domain adaptation effectively reduces inter-domain discrepancies. The synergistic operation of these three components significantly improves the model's diagnostic performance under cross-operating conditions.

Table 5: Ablation experiment results.

ConvNeXt-Transformer	Dual-Signal Masked Contrastive Learning	LMMD + Adversarial Domain Adaptation	Accuracy (%)	ΔAcc (vs. Baseline)
× (CNN-only)	×	×	80.35	—
× (CNN-only)	✓	×	85.89	+5.54
× (CNN-only)	×	✓	90.33	+9.98
× (CNN-only)	✓	✓	94.21	+13.86
✓	×	×	89.88	+9.53
✓	×	✓	95.14	+14.79
✓	✓	×	92.31	+11.96
✓	✓	✓	99.18	+18.83

To further analyze the feature extraction and domain adaptation capabilities of the SSMCL-DA model under variable operating conditions, we performed t-SNE visualizations on the features output from the fully connected layers of each method under the cross-domain transfer task B → C (with a labeled sample size of 80), with the results presented in Fig. 9.

Fig. 9a reveals that the raw data features are chaotically distributed, with samples from different categories overlapping extensively and exhibiting a random pattern. After feature extraction by the respective methods, the resulting feature distributions are shown in Fig. 9b–f. Although the classification performance generally deteriorates compared to the scenario under identical operating conditions, all methods demonstrate a certain degree of clustering tendency. Specifically, the features extracted by the CNN-LSTM, CNN-BiGRU, and CNN-Transformer models still exhibit significant overlap. The features from the MSCNN-LSTM-Attention model show partial overlap across all fault categories. In contrast, the visualization results of the SSMCL-DA model demonstrate superior inter-class separation: only minor confusion exists among a subset of categories (such as between broken tooth, tooth root crack, and missing tooth), while clear separation boundaries are maintained between all other categories. These results indicate that, even in the more challenging transfer scenario of variable operating conditions, SSMCL-DA can still learn highly discriminative and separable features, further validating its comprehensive advantages in both feature extraction and domain adaptation.

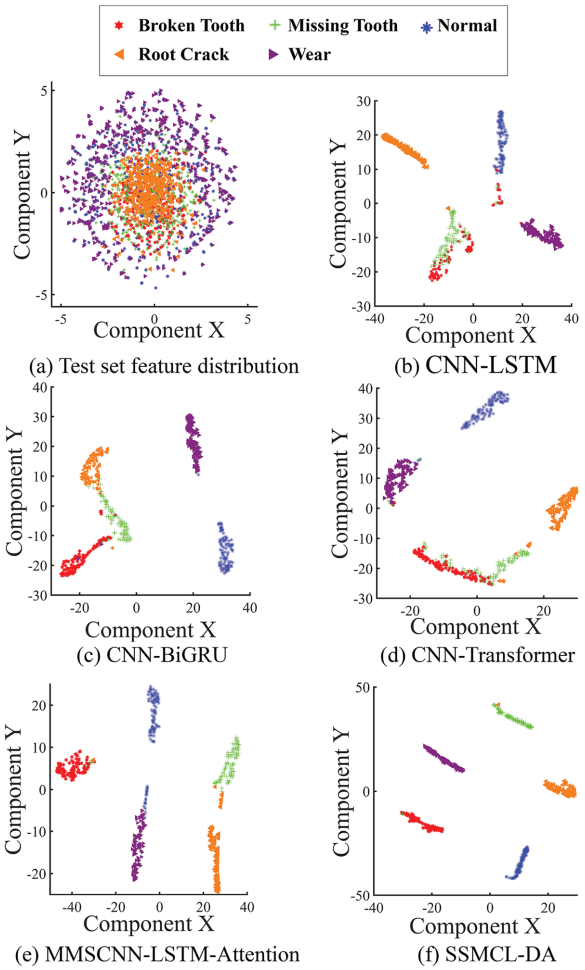

Figure 9: B → C comparison of the feature distribution of each method with a label sample size of 80.

4.5 Bearing Fault Diagnosis under Variable Conditions

To validate the generalization performance of the SSMCL-DA model in cross-equipment fault diagnosis tasks, this study is based on the Case Western Reserve University (CWRU) bearing dataset, which includes data from normal conditions as well as inner race, outer race, and rolling element faults, with a sampling frequency of 12 kHz and load conditions covering 0, 1, 2, and 3 horsepower (HP). Comparative experiments were conducted under variable operating conditions. Using the 3 HP → 0 HP transfer task, the diagnostic accuracy of the SSMCL-DA model was compared with that of the CNN-LSTM and MSCNN-LSTM-Attention models across different labeled sample sizes (ranging from 20 to 100). The results are presented in Table 6. Under all labeled sample size conditions, the diagnostic accuracy of the SSMCL-DA model is significantly higher than that of the two comparative models. Notably, when the labeled sample size increases to 100, the accuracy of SSMCL-DA reaches 99.86%, demonstrating its excellent generalization capability under small-sample conditions.

Table 6: Bearing fault diagnosis accuracy of different models under variable operating conditions (%).

Methods	Number of label samples				
	20	40	60	80	100
CNN-LSTM	85.56	88.15	89.66	90.12	90.25
MSCNN-LSTM-Attention	90.05	93.21	94.51	95.02	95.23
SSMCL-DA	97.32	98.31	98.56	99.02	99.86

To further analyze the feature extraction and domain adaptation capabilities of the SSMCL-DA model in cross-equipment fault diagnosis under variable operating conditions, the t-SNE visualization technique was applied to the features output from the fully connected layers of each method, using the 3 HP → 0 HP transfer task (with 80 labeled samples) as an example. The results are shown in Fig. 10. Fig. 10a shows that the raw feature distribution of the CWRU bearing data under the 0 HP condition is chaotic, with samples from different categories severely overlapping and randomly scattered. The feature distributions extracted by the CNN-LSTM, MSCNN-LSTM-Attention, and SSMCL-DA methods are shown in Fig. 10b–d. It can be observed that the features extracted by the CNN-LSTM model still exhibit significant overlap, resulting in low inter-class discriminability. The feature quality of the MSCNN-LSTM-Attention model shows some improvement, but slight confusion remains between certain categories (e.g., between B007 and B021). In contrast, the SSMCL-DA model demonstrates the best feature separation performance, achieving clear inter-class boundaries across all ten fault conditions (Normal, IR007, B007, OR007, OR014@6, B014, OR014, IR021, B021, OR021) with no observable confusion. These results indicate that even in complex diagnostic scenarios involving cross-equipment and variable operating conditions, the SSMCL-DA model can still learn highly discriminative features that are compact within classes and separable between classes, showcasing its exceptional cross-domain generalization ability and robust fault characterization performance.

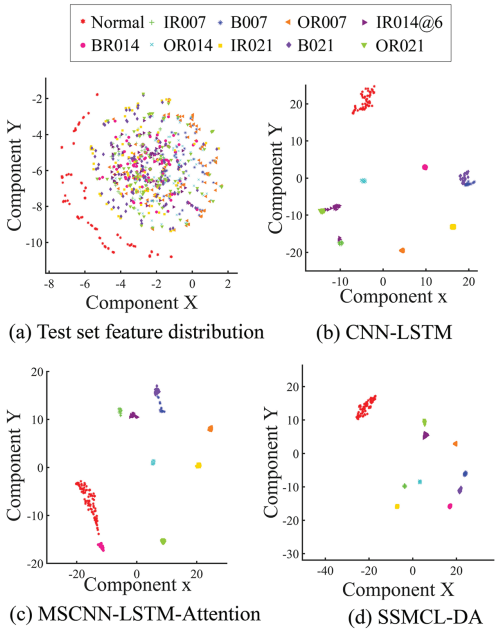

Figure 10: 3 HP → 0 HP comparison of the feature distribution of each method with a label sample size of 80.

5 Conclusion

To address the performance degradation of fault diagnosis models caused by scarce labeled data and distribution shifts under variable operating conditions in gearboxes, this paper proposes a fault diagnosis method integrating Semi-Supervised Masked Contrastive Learning and Domain Adaptation (SSMCL-DA). Through theoretical analysis and experimental validation, the core contributions and conclusions of this research are outlined as follows:

(1) A pre-training strategy based on dual-signal augmented masked contrastive learning is proposed. By simultaneously applying time-domain random masking and frequency-domain random scaling to a large number of unlabeled samples, more discriminative positive sample pairs are constructed. This strategy compels the encoder to learn robust intrinsic feature representations that are insensitive to operational fluctuations under unsupervised conditions, significantly reduces the model's reliance on limited labeled data, and establishes a high-quality initialization for subsequent fine-tuning.

(2) A ConvNeXt-Transformer hybrid encoder is constructed as the backbone feature extraction network. This design effectively integrates the efficiency of the ConvNeXt module in capturing local details with the powerful global dependency modeling capability of the Transformer, thereby enhancing the model's ability for deep representation and extraction of complex fault patterns in vibration signals.

(3) A domain adaptation module combining Local Maximum Mean Discrepancy (LMMD) and adversarial learning is designed. This module achieves fine-grained class-conditional distribution alignment through the LMMD loss and incorporates adversarial training with a domain discriminator to learn globally domain-invariant features. The synergistic operation of these two mechanisms significantly mitigates cross-condition feature distribution discrepancies caused by variations in load, rotational speed, and other factors, thereby improving the model's domain adaptation and generalization performance.

Experimental results demonstrate that the SSMCL-DA model achieves superior performance in variable-condition diagnosis tasks on both the WT planetary gearbox dataset and the CWRU bearing dataset, with peak diagnostic accuracies reaching 99.21% and 99.86%, respectively. The model also exhibits stable generalization capability and robustness in cross-condition and cross-equipment transfer scenarios.

Limitations: The model is somewhat sensitive to the quality of labeled data during the fine-tuning stage. In scenarios with scarce or noisy labels, pseudo-label noise tends to be amplified during iterative processes, leading to residual errors in variable-condition diagnosis across different labeled sample sizes. To enhance the model's practicality and reliability, future work will focus on developing more robust pseudo-label generation and screening mechanisms—such as introducing uncertainty estimation or consistency verification to effectively suppress noise propagation. Concurrently, Wang et al. [27] proposed an enhanced vision-transformer model integrated with semi-supervised transfer learning and successfully conducted transfer experiments across different types of batteries. This provides a strong reference and impetus for us to further extend this method to transfer validation across different types of gearboxes, thereby significantly enhancing the model's generalization capability in complex industrial scenarios.

Acknowledgement: This work was supported by the National Natural Science Foundation of China (Grant No. 52172381). The authors sincerely thank Professor Hanbing Wei for his support and project coordination. We also gratefully acknowledge Zhou Wei from the laboratory for his technical discussions during the experimental process.

Funding Statement: This work was supported by the National Natural Science Foundation of China Funded Project (Project Name: Research on Robust Adaptive Allocation Mechanism of Human Machine Co-Driving System Based on NMS Features; Project Approval Number: 52172381).

Author Contributions: Zhixiang Huang: conceptualization, formal analysis, investigation, methodology, software, validation, visualization, writing—original draft, writing—review & editing, data curation. Jun Li: supervision, investigation, funding acquisition, resources, project administration, writing—review & editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: This study employs two publiche t-distributed Stochastic Neighbor Embedding (t-SNE) ly available benchmark datasets for mechanical fault diagnosis. (1) The WT Planetary Gearbox Dataset: Released by the Key Laboratory of Advanced Manufacturing Technology at Beijing University of Technology, it provides vibration signals from wind turbine planetary gearboxes under various fault conditions, making it applicable for intelligent diagnosis research in complex machinery. The data can be accessed at: https://github.com/Liudd-BJUT/WT-planetary-gearbox-dataset; (2) The Case Western Reserve University Bearing Dataset: Supplied by the Case Western Reserve University Bearing Data Center, it is a classical benchmark dataset containing motor bearing faults in the inner race, outer race, and rolling elements, widely used for model validation. The data is available at: https://engineering.case.edu/bearingdatacenter/apparatus-and-procedures.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xu S, Deng AD, Yang HQ, Fan YS, Deng MQ, Liu DC. Rotating machinery fault diagnosis method based on improved residual neural network. Acta Energiae Solaris Sin. 2023;44(7):409–18. doi:10.19912/j.0254-0096.tynxb.2022-0393.
2. Wang X, Zhang H, Du Z. Multiscale noise reduction attention network for aeroengine bearing fault diagnosis. IEEE Trans Instrum Meas. 2023;72:3513810. doi:10.1109/TIM.2023.3268459.
3. Chen Z, Li C, Sanchez RV. Gearbox fault identification and classification with convolutional neural networks. Shock Vib. 2015;2015(1):390134. doi:10.1155/2015/390134.
4. Hogea E, Onchiş DM, Yan R, Zhou Z. LogicLSTM: logically-driven long short-term memory model for fault diagnosis in gearboxes. J Manuf Syst. 2024;77(11):892–902. doi:10.1016/j.jmsy.2024.10.003.
5. Liu Y, Yu Z, Xie M. Cascading time-frequency transformer and spatio-temporal graph attention network for rotating machinery fault diagnosis. IEEE Trans Instrum Meas. 2024;73:3530310. doi:10.1109/TIM.2024.3453312.
6. Ma Y, Zhang X, Dai W, Zhang C. Intelligent fault diagnosis for offshore wind turbine gearbox: a hybrid framework integrating enhanced VMD and GRU. Int J Electr Power Energy Syst. 2025;173(22):111402. doi:10.1016/j.ijepes.2025.111402.
7. Dutta P, Kanti Podder K, Islam Sumon MS, Chowdhury MEH, Khandakar A, Al-Emadi N, et al. GearFaultNet: novel network for automatic and early detection of gearbox faults. IEEE Access. 2024;12(1):188755–65. doi:10.1109/access.2024.3412274.
8. Yang J, Xu A. A novel fault diagnosis method based on EEWT-EWLTSA and improved deep ELM. Sci Rep. 2025;15(1):41262. doi:10.1038/s41598-025-25203-0.
9. He C, Yasenjiang J, Lv L, Xu L, Lan Z. Gearbox fault diagnosis based on MSCNN-LSTM-CBAM-SE. Sensors. 2024;24(14):4682. doi:10.3390/s24144682.
10. Zhang X, Gu G. Fault diagnosis for rolling bearings under complex working conditions based on domain-conditioned adaptation. Machines. 2024;12(11):787. doi:10.3390/machines12110787.
11. Li T, Zhao Z, Sun C, Yan R, Chen X. Domain adversarial graph convolutional network for fault diagnosis under variable working conditions. IEEE Trans Instrum Meas. 2021;70:3515010. doi:10.1109/TIM.2021.3075016.
12. Nguyen DT, Nguyen VQV, Tran TT, Pham VT. MLFork: bearing fault diagnosis *via* Mamba-powered few-shot learning model with multi-level architecture enhanced by spatial-wise and channel-wise local vector attention. Neurocomputing. 2025;656(1):131518. doi:10.1016/j.neucom.2025.131518.
13. Li DD, Zhao Y, Zhao Y. Fault diagnosis method for wind turbine planetary gearbox under variable working conditions. Electr Mach Control. 2023;27(1):33–45. doi:10.15938/j.emc.2023.01.004.

14. Shao HD, Lin J, Min ZS, Ming YH. Improved semi-supervised prototype network for cross-domain fault diagnosis of gearbox under out-of-distribution interference samples. J Mech Eng. 2024;60(4):212–21. doi:10.3901/jme.2024.04.212.
15. Yu HS, Tang BP, Zhang K, Tan Q, Wei J. Fault diagnosis method of wind turbine gearboxes mixed with attention prototype networks under small samples. China Mech Eng. 2021;32(20):2475–81. doi:10.2139/ssrn.4059122.
16. Lei Z, Zhang P, Chen Y, Feng K, Wen G, Liu Z, et al. Prior knowledge-embedded meta-transfer learning for few-shot fault diagnosis under variable operating conditions. Mech Syst Signal Process. 2023;200(1–2):110491. doi:10.1016/j.ymssp.2023.110491.
17. Tang T, Wang J, Yang T, Qiu C, Zhao J, Chen M, et al. An improved prototypical network with L2 prototype correction for few-shot cross-domain fault diagnosis. Measurement. 2023;217:113065. doi:10.1016/j.measurement.2023.113065.
18. Liang G, Li F, Pang X, Zhang B, Yang P. A gear fault diagnosis method based on reactive power and semi-supervised learning. Meas Sci Technol. 2024;35(12):126107. doi:10.1088/1361-6501/ad71e8.
19. Xu Y, Chen Z, Wang R, Li Y, Tang F, Zhao M, et al. FaultDiffusion: few-shot fault time series generation with diffusion model. arXiv:2511.15174. 2025.
20. He D, He M, Yoon J. Full ceramic bearing fault diagnosis with few-shot learning using GPT-2. Comput Model Eng Sci. 2025;143(2):1955–69. doi:10.32604/cmes.2025.063975.
21. Wang L, Gao Y, Li X, Gao L. Self-supervised pseudo-label learning-enabled cross-domain fault diagnosis method under time-varying speeds. IEEE Trans Syst Man Cybern Syst. 2025;55(9):6203–14. doi:10.1109/TSMC.2025.3576815.
22. Yang X, Yuan X, Ye T, Xu Q, Song Y, Jin J. Cross-layer alignment network with norm constraints for fault diagnosis under varying working conditions. IEEE Trans Instrum Meas. 2023;72:3529513. doi:10.1109/TIM.2023.3312751.
23. Cao H, Shao H, Zhong X, Deng Q, Yang X, Xuan J. Unsupervised domain-share CNN for machine fault transfer diagnosis from steady speeds to time-varying speeds. J Manuf Syst. 2022;62(11):186–98. doi:10.1016/j.jmsy.2021.11.016.
24. Zhou Y, Dong Y, Tang G. Time-varying online transfer learning for intelligent bearing fault diagnosis with incomplete unlabeled target data. IEEE Trans Ind Inform. 2023;19(6):7733–41. doi:10.1109/TII.2022.3230669.
25. Guo J, Chen K, Liu J, Ma Y, Wu J, Wu Y, et al. Bearing fault diagnosis based on deep discriminative adversarial domain adaptation neural networks. Comput Model Eng Sci. 2024;138(3):2619–40. doi:10.32604/cmes.2023.031360.
26. Liu D, Cui L, Cheng W. A review on deep learning in planetary gearbox health state recognition: methods, applications, and dataset publication. Meas Sci Technol. 2024;35(1):012002. doi:10.1088/1361-6501/acf390.
27. Wang YX, Zhao S, Wang S, Ou K, Zhang J. Enhanced vision-transformer integrating with semi-supervised transfer learning for state of health and remaining useful life estimation of lithium-ion batteries. Energy AI. 2024;17(1–2):100405. doi:10.1016/j.egyai.2024.100405.