machines

MDPI

Article

Semi-Supervised Fault Diagnosis Method for Hydraulic Pumps Based on Data Augmentation Consistency Regularization

Siyuan Liu [1,2], Jixiong Yin [1,2], Zhengming Zhang [3,*], Yongqiang Zhang [1,2], Chao Ai [1,2] and Wanlu Jiang [1,2]

[1] Hebei Provincial Key Laboratory of Heavy Machinery Fluid Power Transmission and Control, Yanshan University, Qinhuangdao 066004, China; liusiyuan@ysu.edu.cn (S.L.); 1462910998yjx@stumail.ysu.edu.cn (J.Y.); zyq@stumail.ysu.edu.cn (Y.Z.); aichao@ysu.edu.cn (C.A.); wljiang@ysu.edu.cn (W.J.)
[2] State Key Laboratory of Crane Technology, Yanshan University, Qinhuangdao 066004, China
[3] Aviation Maintenance NCO School, Air Force Engineering University, Xinyang 464000, China
* Correspondence: zzm_zhang1202@163.com

Abstract

Due to the scarcity of labeled samples, the practical engineering application of deep learning-based hydraulic pump fault diagnosis methods is extremely challenging. This study proposes a semi-supervised learning method based on data augmented consistency regularization (DACR) to address the issue of lack of labeled data in diagnostic models. It utilizes augmented data obtained from the improved symplectic geometry modal decomposition method as additional perturbations, expanding the feature space of limited labeled samples under different operating conditions of the pump. A high-confidence label prediction process is formulated through a threshold determination strategy to estimate the potential label distribution of unlabeled samples. Consistent regularization loss is introduced in labeled and unlabeled data, respectively, to regularize model training, reducing the sensitivity of the classifier to additional perturbations. The supervised loss term ensures that the predictions of the augmented labeled samples are consistent with the true labels. Meanwhile, the unsupervised loss term can be used to minimize the difference between the distributions of unlabeled samples for different augmented versions. Finally, the proposed method is combined with Kolmogorov–Arnold Network (KAN). Comparative experiments based on data from two models of hydraulic pumps verify the superior recognition performance of this method under low label rate.

Keywords: fault diagnosis; hydraulic pumps; semi-supervised learning; data augmentation; deep learning

check for updates

Academic Editor: Ahmed Abu-Siada

Received: 20 May 2025
Revised: 20 June 2025
Accepted: 25 June 2025
Published: 26 June 2025

Citation: Liu, S.; Yin, J.; Zhang, Z.; Zhang, Y.; Ai, C.; Jiang, W. Semi-Supervised Fault Diagnosis Method for Hydraulic Pumps Based on Data Augmentation Consistency Regularization. *Machines* **2025**, *13*, 557. https://doi.org/10.3390/machines13070557

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The hydraulic pump is the core power element of the heavy equipment hydraulic system [1–3]. Its quality is directly related to the safe and stable operation of the whole system. Once a failure occurs, it may cause equipment downtime, or bring serious safety accidents [4,5]. Therefore, an accurate fault diagnosis of hydraulic pumps under high load conditions not only helps to prevent potential risks but also is an important means of guaranteeing their long-term stable operation in harsh environments [6,7]. Intelligent fault diagnosis methods have received more attention in recent years, but the diagnostic models usually require a large number of labeled samples during the training process, which limits the realistic application of the method [8,9]. The closed nature of hydraulic pumps causes many uncertainties in the occurrence of failures, which are hidden and difficult to detect,

making it exceptionally difficult to obtain high-quality labeled samples [10–12]. In contrast, unlabeled data are more common and easily accessible in practical engineering [13–15]. This data imbalance between the scarcity of labeled data and the abundance of unlabeled data has prompted scholars to explore solutions to improve model performance, which has now become a popular issue in the field of intelligent fault diagnosis [16].

In recent years, semi-supervised learning (SSL) has gradually gained popularity in the field of fault diagnosis [17–19]. Especially when facing the problem of scarcity of labeled data, semi-supervised learning methods are able to utilize both limited labeled data and a large amount of unlabeled data to alleviate the problem of difficulty in obtaining insufficient labeled data to a certain extent [20,21]. This approach not only improves the learning efficiency of the model, but also significantly improves the accuracy of fault diagnosis with extremely limited labeled data. He et al. [22] proposed an encoder network with a two-channel heterogeneous convolution kernel to extract fault features from a small number of samples while using a similarity function to identify unlabeled data to fine-tune the network for fault classification under the scarcity of labeled data. However, the model exhibits volatility on different datasets, leading to insufficient generalization ability. Han et al. [15] used adversarial learning to construct a semi-supervised deep neural network in order to cope with the problem of the scarcity of annotated samples for rotating machinery and employed metric learning-guided discriminant feature enhancement techniques to improve the separability of different manifolds. However, the appropriate metric function is crucial to the model performance impact, while the noise samples have a large impact on the metric learning model. Ozdemir et al. [23] presented a semi-supervised approach based on the student–teacher model, which utilizes the information from the pre-trained model to label a lot of unlabeled data and uses the pseudo-labeled data for the model training procedure to decrease the workload of manual labeling process. However, incorrect pseudo-labeling is highly likely to make the model difficult to converge. Azar et al. [24] developed a novel hybrid semi-supervised fault diagnosis model incorporating a large amount of monitoring data by combining statistical learning methods, which is further combined with reinforcement learning optimization strategies to reduce the dependence on data a priori knowledge and assumptions. However, a large number of different statistical features of the data may lead to a significant reduction in model solving efficiency. In addition to the above methods, many transfer learning strategies exist for realizing semi-supervised learning tasks. The core of these strategies is to migrate the rich labeled knowledge in the source domain to the target domain which lacks labeled data. Su et al. [25] developed a deep semi-supervised transfer learning strategy aimed at achieving sensitivity-aware adaptive decision borders. The approach works by decreasing the discrepancy between the prediction matrices of source and target domain data in order to allow the decision border to adapt more efficiently to target domain data. Lu et al. [26] designed a deep directed migration network that incorporates the mechanism of clustering pseudo-label learning. The goal of the model is to effectively decrease the distance between features in each subdomain by controlling the minimization and maximization of feature entropy together with the amount of linearly separate vectors in the target domain. Kumar et al. [27] proposed a novel multi-domain learning network based on semi-supervised transfer learning, using a combination of encoder network and attention mechanism for solving the problem of the lack of labeled data during training. Existing semi-supervised fault diagnosis techniques with transfer learning strategies show that it can alleviate the problem of scarcity of labeled samples in the target domain to some extent. However, constructing such a model is extremely dependent on sufficient labeled samples in the source domain, and the problem of limited device labeled samples remains unresolved in practical applications.

In order to solve the problem of the lack of a sufficient number of labeled samples in semi-supervised learning, an effective approach is to generate new data with similar feature distributions as the original samples through data enhancement techniques. The core of this approach lies in transforming or expanding the original samples to generate new samples so that the model can see a more diverse distribution of data during the training process, thus improving the generalization ability of the model. By effectively utilizing these augmented data, the lack of labeled samples can be compensated to a certain extent, and this approach coincides with the principle of anti-perturbation in semi-supervised consistency methods. Yu et al. [28] combined a re-parameterized residual feature network with a de-noising diffusion probabilistic model to generate high-quality signal samples through a process of forward diffusion and backward de-noising for fault diagnosis tasks in rotating machinery. Kulevome et al. [29] proposed an innovative analytical wavelet data augmentation method to synthesize a sample of scale maps close to the properties of the original samples by adjusting the parameters of the generalized morse wavelet. Tian et al. [30] incorporated a learning algorithm with adaptive loss in a variational auto-encoder to alleviate the widespread problem of Kullback–Leibler scatter vanishing during training of the generator. Mueller et al. [31] added an attention mechanism to the diffusion model for generating data samples from different state categories. Most of the additional perturbations in the traditional data enhancement methods mentioned above are designed for 2D images, while the monitoring data of pump rotating machinery are usually 1D time series signals. Therefore, these perturbations cannot be directly utilized to achieve the fault diagnosis of hydraulic pumps.

Considering the above problems and inspired by the examples of semi-supervised learning and data augmentation methods in other rotating machinery fault diagnosis outcomes, this research presents a semi-supervised learning approach with improved symplectic geometry data augmentation with consistency regularization, which solves the problem of insufficient generalization ability of the supervised learning model under the scarcity of labeled data. The new data consistent with the feature distribution of the original samples is generated, which enriches the feature space of the labeled samples of the pump under different operating conditions. A consistent regularization loss for both labeled and unlabeled data is introduced to enhance the robustness of the model. The consistency of the predictions of the augmented labeled samples with the actual labels is ensured by supervised loss, while the distributional differences between the different augmented samples and the unlabeled samples are reduced by unsupervised loss. Finally, the proposed method is validated in conjunction with KAN on two datasets of hydraulic pumps with different displacements. The t-distributed Stochastic Neighbor Embedding (T-SNE) visualization is also introduced to further analyze the classification effect of each model on data features [32]. The analysis results show that the method proposed in this paper has superior fault diagnosis performance compared to other comparative methods in the presence of extremely sparse labeled data.

The main contributions of this research are as follows:

(1) A semi-supervised learning method based on data augmentation and consistency regularization is proposed. Utilizing the improved symplectic geometry data augmentation approach (ISGDA), the amount of labeled samples is enriched by obtaining augmented samples by applying additional perturbations to the temporal sequence signals of 1D failure samples. The results of fault diagnosis test trials indicate that the ISGDA dramatically enhances the diagnostic effect of the model in situations where the labeled failure data is rare, and meanwhile effectively suppresses the overfitting problem in the training.

(2) The consistency strategy among primitive labeled samples and enhanced samples is constructed, and the supervised loss function is defined. Standard cross entropy is calculated for enhanced labeled samples to effectively improve the classification performance of the semi-supervised task under the condition that label info of the marked samples is always kept constant.

(3) A prediction mechanism is designed to discriminate the potential label distribution of unlabeled samples after augmentation, and an unsupervised consistency loss function is constructed in order to minimize the distributional gap among unlabeled augmented samples.

The remainder of the article is structured as below. Section 2 illustrates the fundamentals of the proposal approach. Section 3 describes the general modeling framework of DACR semi-supervised approach. Section 4 demonstrates the test results for two different displacement hydraulic pumps. Section 5 gives the conclusion of the research.

2. Basic Theory

2.1. Symplectic Geometry Modal Decomposition

Currently, the popular signal decomposition methods in the field of fault diagnosis include empirical mode decomposition [33], local characteristic decomposition [34], singular spectrum analysis [35], etc. As a rising star, the symplectic geometry modal decomposition method (SGMD) uses symplectic geometry similarity transformation compared to commonly used methods. Its advantages are that it keeps the essential characteristics of the original time series unchanged and can effectively suppress modal aliasing. It has obvious advantages without changing the essential characteristics of the time series and without the need for custom parameters [36,37]. Figure 1 summarizes the flowchart of the SGMD, which can be divided into the following four steps:

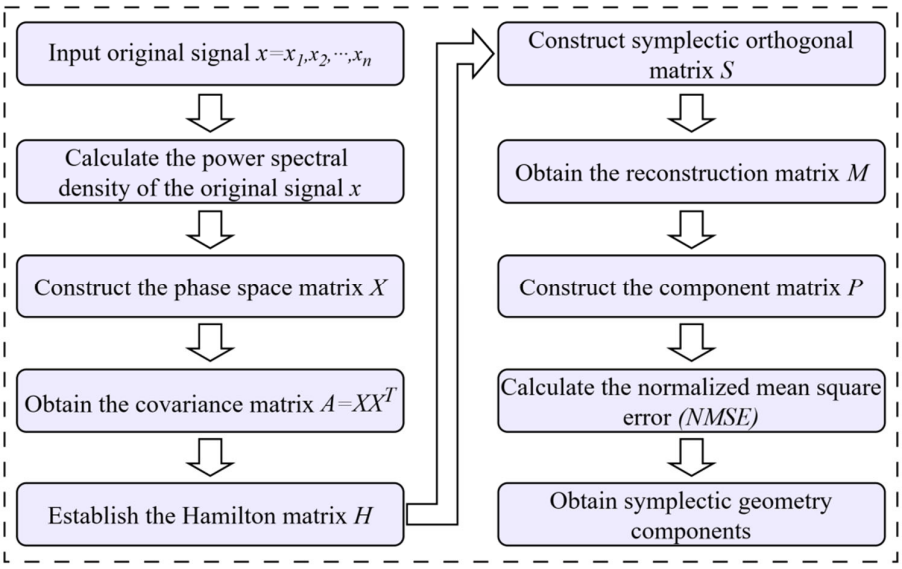


Figure 1. The flowchart of SGMD.

(1) Dynamic selection of insertion dimensions based on signal characteristics.

Based on the raw signal $x = x(x_1, x_2, \cdots x_n)$, a phase space matrix X is generated as follows:

$$X = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(d-1)\tau} \\ \vdots & & \ddots & \vdots \\ x_m & x_{m+\tau} & \cdots & x_{m+(d-1)\tau} \end{bmatrix} \quad (1)$$

In the above equation, d denotes the insertion dimension, τ is delay time, $m = n - (d-1)\tau$, and n is raw signal length.

(2) Compute the characteristic values of the Hamiltonian matrix.

Establish the Hamiltonian matrix H utilizing the phase space matrix X as follows:

$$H = \begin{bmatrix} A & 0 \\ 0 & -A^T \end{bmatrix} \tag{2}$$

where $A = XX^T$.

Reorganization of the symplectic orthogonal matrix S,

$$S^T H^2 S = \begin{bmatrix} B & R \\ 0 & B^T \end{bmatrix} \tag{3}$$

where B is the up-triangular matrix.

Compute the reorganization matrix M as follows:

$$\begin{cases} Y_i = S^T X^T \\ M_i = SY_i \\ M = M_1 + M_2 + \cdots + M_i \end{cases} \tag{4}$$

(3) Average of diagonal elements.

Calculate the average of the diagonal elements of the matrix M to construct the signal fraction matrix P as follows:

$$P_k = \begin{cases} \dfrac{1}{k} \sum\limits_{p=1}^{k} M^*_{p,k-p+1} \; 1 \le k \le d^* \\ \dfrac{1}{d^*} \sum\limits_{p=1}^{d^*} M^*_{p,k-p+1} \; d^* \le k \le m^* \\ \dfrac{1}{n_1 - k + 1} \sum\limits_{p=k-m^*+1}^{n-m^*+1} M^*_{p,k-p+1} \; m^* < k \le n_1 \end{cases} \tag{5}$$

$$P = P_1 + P_2 + \cdots + P_k \tag{6}$$

where $d^* = min(m, d)$, $m^* = max(m, d)$, $n_1 = m + (d-1)\tau$. If $m < d$, then $M^*_{ij} = M_{ij}$, otherwise $M^*_{ij} = M_{ji}$.

(4) Dynamic reorganization of signaling fractions.

The reconfiguration fraction SGC^n is derived by adding up the strongly similar signal fractions and further finding the normalized mean square error of the raw signal x with respect to the decomposed remnant signal g^h. It is specified as follows:

$$NMSE^h = \frac{\sum_{e=1}^{n} g^h(e)}{\sum_{e=1}^{n} x(e)} \tag{7}$$

where h represents the iteration number.

A threshold value of 1% is set as the determination value of the standardized average square error, and when the mistake is greater than the set threshold value, the residual matrix is set as the new primitive matrix, and the loop iteration is carried out. The decomposition is finished when the mistake is less than the set threshold. The details are as follows:

$$x(n) = \sum_{h=1}^{N} SGC^h(n) + g^{(N+1)}(n) \tag{8}$$

where N is quantity of fractions.

2.2. Kolmogorov-Arnold Network

KAN has learnable activation functions at the edges as compared to traditional Multi-Layer Perceptron Networks. Meanwhile, the use of spline to represent the weights improves the ability to approximate complex functions with fewer parameters. KAN is inspired by the Kolmogorov–Arnold representation theorem [38], that states any multivariate sequential function is realized by a finite number of single-variable functions and combinations of additive operations. Formally, for a smooth function $f : [0,1]^n \to \mathbb{R}$, this can be expressed as:

$$f(x) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \varphi_{q,p}(x_p) \right) \tag{9}$$

where $\varphi_{q,p} : [0,1] \to \mathbb{R}$ and $\Phi_q : \mathbb{R} \to \mathbb{R}$ are continuous functions.

In KAN, weight parameters are replaced by learnable 1D functions φ, parametrized as B-splines. The computation in a KAN layer with n_{in} inputs and n_{out} outputs is:

$$x_{l+1,j} = \sum_{i=1}^{n_l} \varphi_{l,j,i}(x_{l,i}) \tag{10}$$

where $\varphi_{l,j,i}$ is a spline function connecting the i-th neuron in layer l to the j-th neuron in layer $l+1$.

The backpropagation process in KAN involves calculating gradients of the spline functions. The loss $\mathcal{L}$ is minimized using gradient descent, with the gradient of the loss with respect to the spline parameters c_i computed as:

$$\frac{\partial \mathcal{L}}{\partial c_i} = \sum_{j=1}^{n_{out}} \frac{\partial \mathcal{L}}{\partial x_{l+1,j}} \frac{\partial x_{l+1,j}}{\partial c_i} \tag{11}$$

where $\frac{\partial x_{l+1,j}}{\partial c_i}$ involves the derivative of the spline function with respect to its coefficients.

3. The Overall Methodological Framework

The aim of this research is to combine finite labeled data with a lot of unlabeled data to increase the classification capability of the model for more accurate hydraulic pump troubleshooting. Figure 2 indicates that only very few samples in the original dataset have label information, and most of the samples are in the unlabeled state. Using the traditional supervised model to train the data is very easy to produce wrong decision boundaries, resulting in poor diagnostic results. This paper utilizes the ISGDA method to generate different enhanced versions of data with similar feature distributions as the original samples. The respective consistency regularization loss is constructed for both labeled and unlabeled data, aiming to enhance the predictive performance of the model for unlabeled data, and thus optimize the decision boundary for best classification results.

3.1. A Data Augmentation Approach Based on Improved Symplectic Geometry

This paper constructs a consistency strategy based on the symplectic geometry modal decomposition method concerning the raw data and augmented samples with average and normal deviation, which applies additional perturbations to the samples while preserving the effective fault characteristics. Four different data enhancement methods including amplitude scaling, overall flipping, local slice flipping and adding Gaussian white noise are proposed. The ISGDA specific methodology flow is shown in Figure 3.

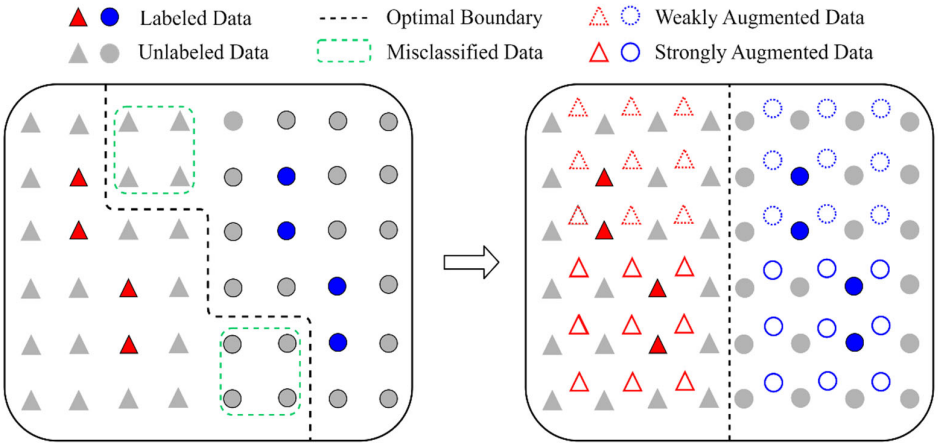


Figure 2. Diagram of the principle of the proposed approach.

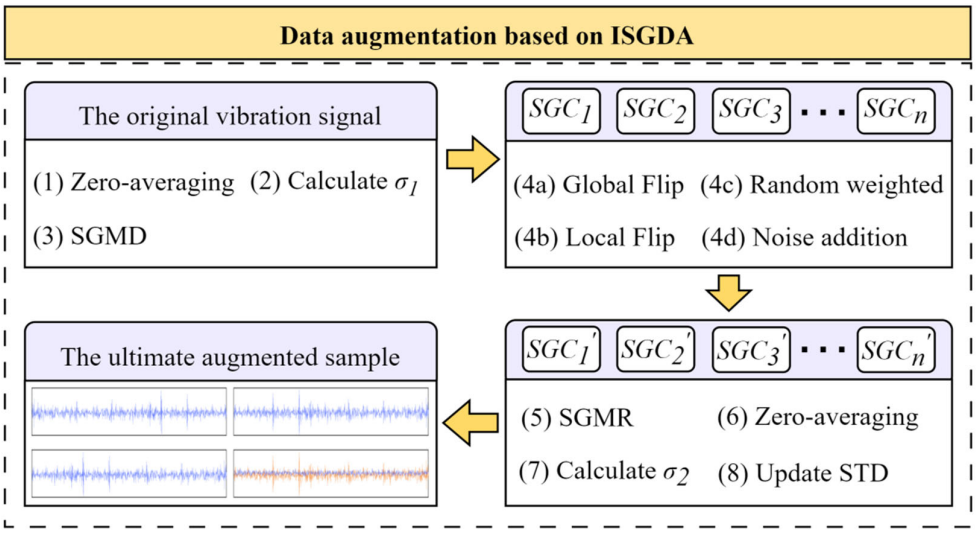


Figure 3. The flowchart of the ISGDA method.

(1) The symplectic geometry mode decomposition.

Set the raw time series signal to $x(t)$ and perform zero-averaging as follows:

$$x'(t) = x(t) - \mu(x) \tag{12}$$

where $x(t) = [x_1, x_2, \ldots, x_N]$, $\mu(x) = (x_1 + x_2 + \cdots + x_N)/N$, $x'(t)$ as a sign of decentralization.

The variance σ_1 of the raw time series signal is solved as follows:

$$\sigma_1 = \frac{\sum(x_i - \mu(x))^2}{N} \tag{13}$$

where $i = 1, 2, \ldots, N$.

Adaptive decomposition of the raw signal to various symplectic geometry components is performed by the SGMD as follows [39]:

$$P = P_1 + P_2 + \cdots + P_k \tag{14}$$

where P is the component matrix.

(2) The four augmentation strategies.

a. Overall flip:

Inspired by the flip operation in image transformation, a component is randomly selected, and the selected component is flipped along the time dimension. The result of this strategy is shown in Figure 4a, as shown in the following Equation (15).

$$P'_x = reverse(P_x) \tag{15}$$

where P_x is the selected fraction, P'_x is the enhanced fraction, $x = 1, 2, \ldots, k$.

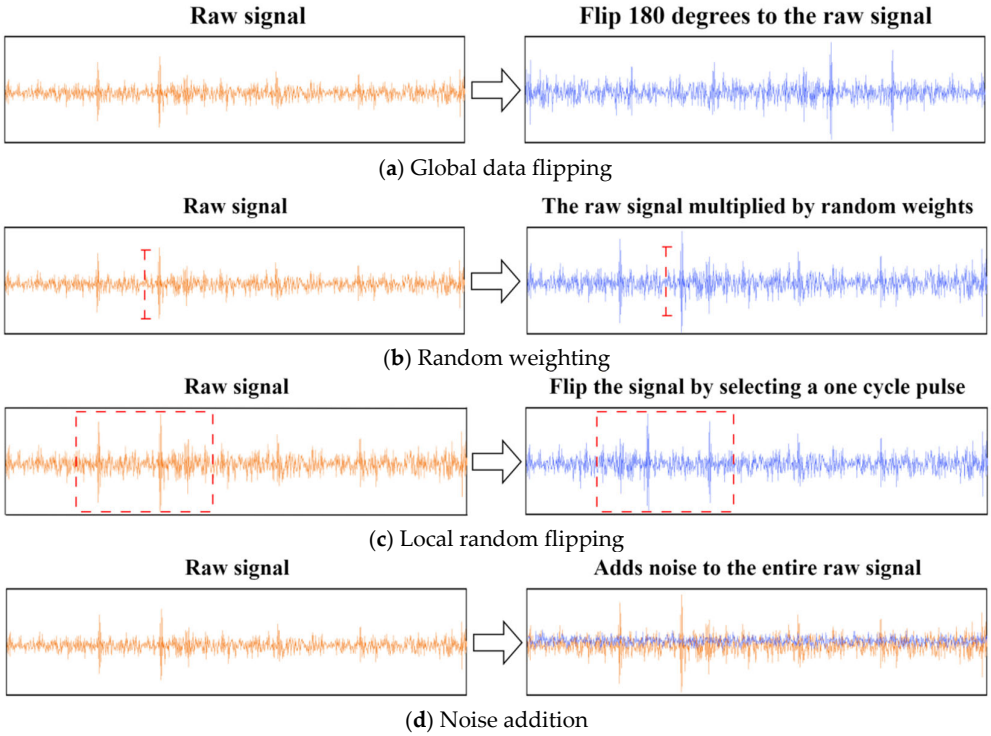


Figure 4. A schematic diagram of different data enhancement strategies.

b. Random weighting:

The random selection of a certain component is weighted, and the optimal preset weight range is [0.6,1.8] in this paper through the data distribution and model parameter experiments. The raw data is multiplied with the weights within a randomly selected threshold to obtain the enhanced fraction. This strategy results in Figure 4b, specifically in Equation (16) below.

$$P'_x = \alpha P_x \tag{16}$$

where α is the weight.

c. Partial flipping:

A component is randomly selected from which a local data segment of length one cycle pulse signal is randomly selected, and the segment is flipped along the time dimension. The result of this strategy is shown in Figure 4c, as shown in the following Equation (17).

$$\begin{cases} M = f_s \times T \\ P_l = \{x_j, x_{j+1}, \cdots, x_{j+M-1}\}, \ 1 \le j \le P_x - M + 1 \\ P'_x = reverse(P_l) \end{cases} \tag{17}$$

where M is one cycle of sampling dots, f_s is the sampling frequency, T is the time of a cycle, j is the randomly selected starting position, and P_l is the randomly selected local data segment.

d. Randomly added white noise:

Gaussian white noise can be viewed as an additional perturbation over the entire length of the data, and classification models tend to be insensitive to additional perturbations as a way to better improve their generalization. A component is randomly selected and a Gaussian white noise with a signal-to-noise ratio (SNR) of 20 dB is added, and the length of the noise is equal to the length of the selected component. The result of this strategy is shown in Figure 4d, as shown in Equation (18) below.

$$P'_x = P_x + \epsilon,\ \epsilon \ \sim \ N(0, \sigma_n) \tag{18}$$

where ϵ is the noise, σ_n is the variance of the noise.

(3) Augmented Signal Reconstruction.

The signal components after adding the perturbation augmentation are reorganized with the residual fractions to synthesize the novel vibration signal $y(t)$ as follows:

$$y(t) = P_1 + P'_x + \cdots + P_k \tag{19}$$

Zero-means the restructured data, that is:

$$y'(t) = y(t) - \mu(y) \tag{20}$$

where $\mu(y) = (P_1 + P'_x + \cdots + P_k)/k$, $y'(t)$ is the zero-mean signal.

The variance σ_2 is computed for the reorganized signal as follows:

$$\sigma_2 = \frac{\sum(P_j - \mu(y))^2}{k} \tag{21}$$

where $j = 1, x, \ldots, k$.

The variance of the restructured signal and the raw signal is calculated and adjusted to match as follows:

$$z(t) = y(t) \cdot \frac{\sigma_1}{\sigma_2} \tag{22}$$

where $z(t)$ is the final enhanced sample.

The algorithm proposed in this paper performs two types of augmentation for the original data, weak and strong augmentation, denoted by $ISGDA_{weak}(\cdot)$ and $ISGDA_{strong}(\cdot)$, respectively. In all experiments, weak enhancement uses only the first enhancement strategy, while strong enhancement is randomly selected among the remaining strategies with uniform probability.

3.2. The Objective Function

The objective function of the proposed method in this paper mainly consists of two cross-entropy loss terms, supervised consistency loss and unsupervised consistency loss.

a. Supervised consistency loss: To ensure that the prediction results of the augmented samples after the addition of perturbations are consistent with the true labels, by optimizing the loss of cross-entropy among the prediction results of the true labels and those of the weakly augmented samples, it can reduce the impact of the additional perturbations and enable the model to have a stable base-learning capability.

Assume that the labeled dataset is denoted as $D_{labeled} = \{(x_i, y_i)\}_{i=1}^{N_l}$, where x_i denotes the i-th labeled sample, y_i denotes the true label of the i-th labeled sample, and N_l denotes the number of samples in the labeled dataset. The unlabeled dataset is denoted as $D_{unlabeled} = \{x_j\}_{j=1}^{N_u}$, where x_j denotes the j-th unlabeled sample and N_u denotes the

number of samples in the unlabeled dataset. Supervised loss is introduced ℓ_{sup}, as shown in the following Equation (23).

$$\ell_{sup} = \frac{1}{B}\sum_{i=1}^{B} H(y_i, p_i) \tag{23}$$

where p_i is the predicted probability distribution of the model for the i-th weakly augmented labeled sample, H is the cross-entropy loss for the i-th sample, which represents the difference between the true label y_i and the predicted distribution p_i, and B is the batch size of the labeled data.

b. Unsupervised consistency loss: Generate high-confidence samples through the label prediction mechanism, so that the model's prediction of weakly enhanced unlabeled data is entropy-minimizing. To generate pseudo-labels for the weakly enhanced unlabeled samples with "one-hot" probability distribution, as shown in Equation (24) below.

$$\begin{cases} \frac{1}{\mu B}\sum_{i=1}^{\mu B} \mathbb{I}(max(p_{iu}) \geq \tau) \cdot H(\hat{y}_i, p_{iu}) \\ \hat{y}_i = argmax(p_{iu}) \end{cases} \tag{24}$$

where μ is the proportion of unlabeled data, μB is the batch size of unlabeled data, p_{iu} is the predicted probability distribution of the model after weakly augmenting the unlabeled data, and τ is the threshold value. This paper sets a threshold value of 0.95, which means that pseudo-labels with model prediction probability higher than the threshold value of 0.95 are retained.

The next step is to minimize the cross-entropy loss among the prediction results of pseudo-labeled and strongly enhanced unlabeled samples, forcing the model to make consistent predictions for the same unlabeled samples under different augmented perspectives, and improving the model generalization ability and overall performance. Unsupervised loss ℓ_{unsup} is introduced, as shown in the following Equation (25).

$$\ell_{unsup} = \frac{1}{\mu B}\sum_{i=1}^{\mu B} \mathbb{I}(max(p_{iu}) \geq \tau) \cdot H(\hat{y}_i, q_i) \tag{25}$$

where q_i is the predictive probability distribution of the model after strong augmentation for unlabeled samples.

Combining the above supervised and unsupervised consistency losses, the two loss terms are combined, and the overall optimization function is shown in Equation (26) below.

$$\ell = \ell_{sup} + \lambda_u \ell_{unsup} \tag{26}$$

where λ_u is a fixed scalar hyper-parameter indicating the relative weight of the unlabeled loss, which is set to 1 in this paper.

3.3. Overall Modeling Framework for Fault Diagnosis

To address the challenge of scarcity of labeled samples in hydraulic pump troubleshooting, this research presents a semi-supervised learning approach using DACR, whose pseudo-code is detailed in Algorithm 1. In this method, an innovative approach to data augmentation for symplectic geometry reconstruction that incorporates multiple augmentation strategies is proposed. Corresponding loss of consistency regularization mechanisms are designed for labeled and unlabeled data, respectively. By introducing cross-entropy loss, it ensures that the enhanced labeled samples accurately match their

true labels. Meanwhile, unsupervised loss focuses on reducing the distributional bias of unlabeled samples among different enhanced versions.

Algorithm 1. The pseudo-code for DACR approach

1: **Input:** Labeled dataset $D_{labeled} = \{(x_i, y_i)\}_{i=1}^{N_l}$; unlabeled dataset $D_{unlabeled} = \{x_j\}_{j=1}^{N_u}$; confidence threshold τ; unlabeled data ratio μ; unlabeled loss weight λ_u; the maximum iterations *epoch*; batch size B.
2: Initialize the network model parameters.
3: Weak enhancement for labeled data.
$\{(\hat{x}_i)\}_{i=1}^{N_l} = ISGDA_{weak}\{(x_i)\}_{i=1}^{N_l}$.
4: Weak and strong enhancement of unlabeled data.
$\{(\hat{x}_j)\}_{j=1}^{N_u} = ISGDA_{weak}\{(x_j)\}_{j=1}^{N_u}$, $\{(\hat{x}_j)\}_{j=1}^{N_u} = ISGDA_{strong}\{(x_j)\}_{j=1}^{N_u}$.
5: **for** *epoch* = 1 **to** *epoch* **do**.
6: **for** B = 1 **to** B **do**.
7: Cross-entropy loss for labeled data $\ell_{sup} = \frac{1}{B}\sum_{i=1}^{B} H(y_i, p_i)$.
8: **for** B = 1 **to** μB **do**.
9: Weakly enhanced label prediction for unlabeled data.
$\frac{1}{\mu B}\sum_{i=1}^{\mu B} \mathbb{I}(max(p_{iu}) \geq \tau) \cdot H(\hat{y}_i, p_{iu})$, $\hat{y}_i = argmax(p_{iu})$.
10: **end for**
11: Cross-entropy loss for pseudo-label and strongly enhanced prediction results
$\ell_{unsup} = \frac{1}{\mu B}\sum_{i=1}^{\mu B} \mathbb{I}(max(p_{iu}) \geq \tau) \cdot H(\hat{y}_i, q_i)$.
12: Calculate $\ell = \ell_{sup} + \lambda_u \ell_{unsup}$.
13: Calculate and update network model parameters.
14: **end for.**
15: **end for.**
16: **Return** The trained network model.

In this paper, the KAN is combined to perform analysis and validation in Figure 5. The trial of the hydraulic pump in various working situations is firstly carried out to obtain the vibration signal and perform data preprocessing. Next, the suggested ISGDA is employed on the delineated dataset to generate different enhanced versions of the data, matching the features of the original samples, enriching the feature space of the pump with finitely labeled examples in various operating states. Then, the model training process is normalized by supervised and unsupervised consistency loss to enhance the model's anti-perturbation ability, which is combined with the KAN. Finally, various types of results from the model diagnostic analysis are visualized.

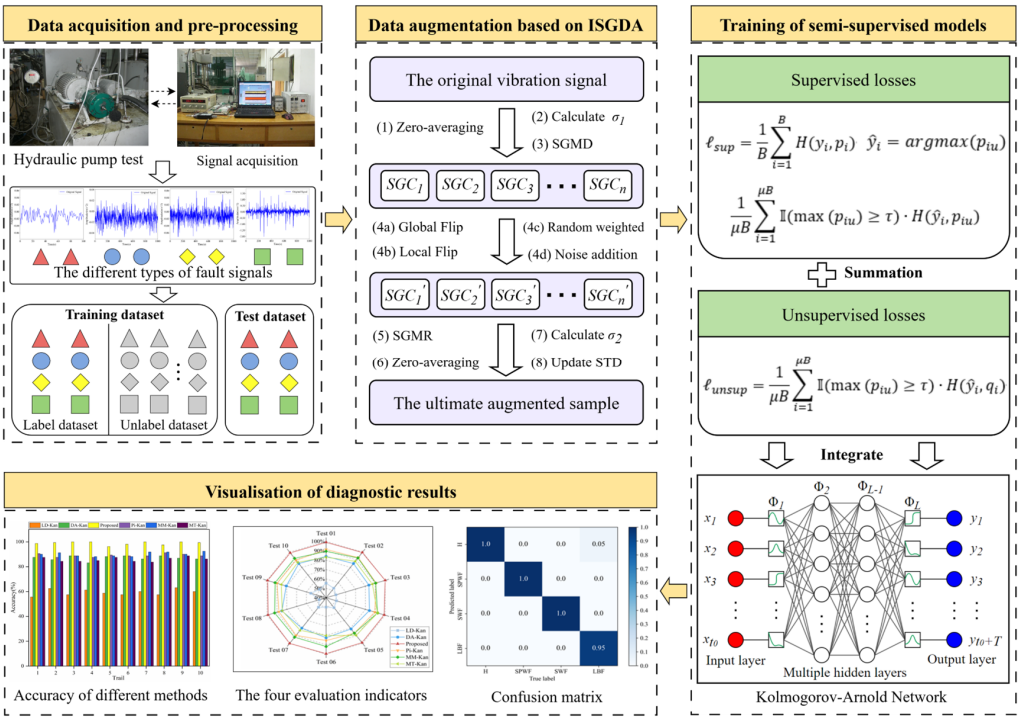


Figure 5. The overall framework of DACR semi-supervised fault diagnosis methodology.

4. Experimental Analysis

This section validates the effectiveness of the proposed semi-supervised approach based on data augmentation and consistency regularization for hydraulic pump fault diagnosis with limited labeled samples through two test case studies. The specific performance parameters are shown in Table 1, and the data used in the two cases are from the hydraulic pump failure test datasets from different test benches of the subject group.

Table 1. Description of experimental data.

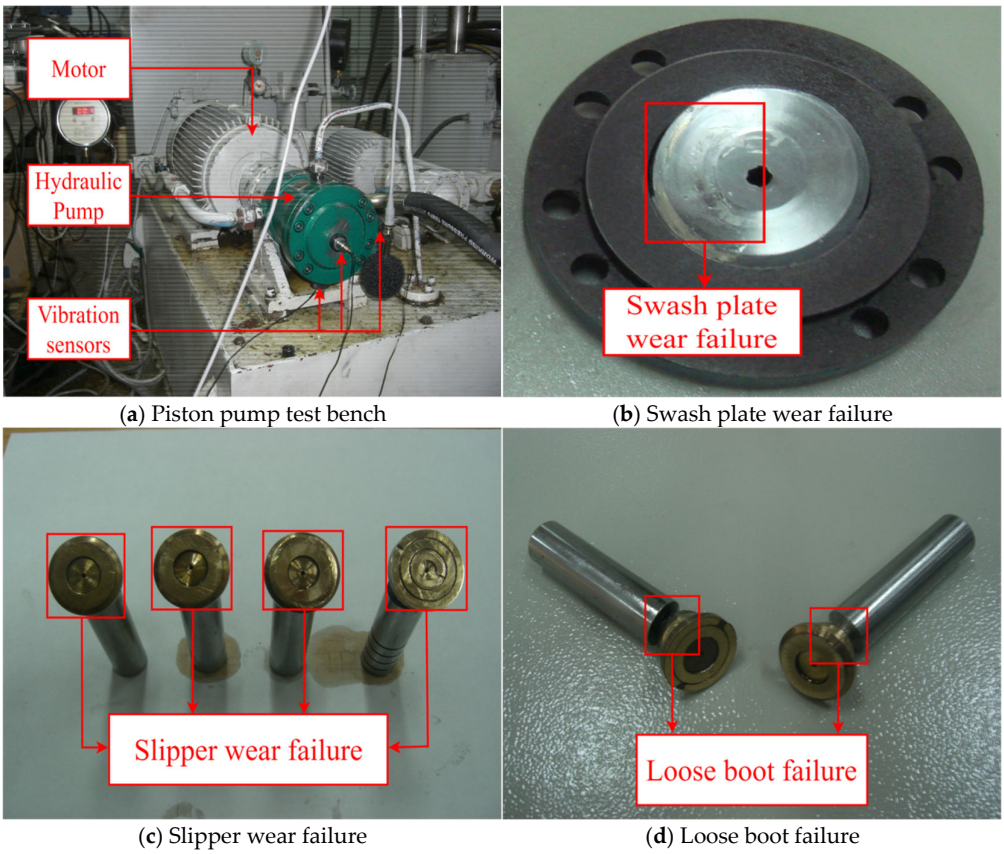
Case	Hydraulic Pump Models	Rated Pressure (Mpa)	Rated Displacement (ml/r)	Rated Speed (r/min)	Number of Plungers	Weight (Kg)
1	10MCY14-1B	31.5	10	1500	7	16.4
2	P08-B3F-R-01	21.5	8	1450	9	15.6

4.1. Case 1: Type 10MCY14-1B Fault Emulation Test Platform

(1) Overview of the test system and data.

The 10MCY14-1B fault simulation platform consists of swashplate axial plunger pumps, acceleration sensors, AC motors, and other components in Figure 6. Four hydraulic piston pump working situations are simulated separately: normal condition, swash plate abrasion, sliding shoe abrasion and sliding shoe loosening. The motor rotation frequency is constant at 1500 rpm and the sampling frequency is 10 kHz with a data duration of 10 s. As shown in Table 2, there are 199 experimental data points for each state. Using a stratified random division method, the training samples and test samples for each state are divided in an 8:2 ratio to ensure that the distribution of each state in the subsets is consistent [40,41]. In order to restore the scarcity of labeled samples in the real state, only (5% of the training samples) 8 out of the 159 training samples in each state are labeled, and the rest are treated as unlabeled samples. In addition, experiments comparing performance

under many different labeled sample ratios are performed in Section 4.3. Robustness experiments under different noise conditions are performed in Section 4.4.



(**a**) Piston pump test bench (**b**) Swash plate wear failure

(**c**) Slipper wear failure (**d**) Loose boot failure

Figure 6. Test rigs for hydraulic piston pumps and faulty components.

Table 2. Case 1: Statement of test data.

Label	Healthy Conditions	Description	Number of Training Datasets	Number of Test Datasets
0	H	State of health	159	40
1	SPWF	Swash plate wear failure	159	40
2	SWF	Slipper wear failure	159	40
3	LBF	Loose boot failure	159	40

The constructed semi-supervised fault diagnosis framework using DACR has been achieved in Python 3.8 setting. By a range of contrasting trials, it has been demonstrated that the method is able to effectively carry out fault diagnosis of pump rotating machinery in the presence of a scarcity of labeled samples.

As a novel neural network architecture, KAN has excellent feature acquisition capability and classification property, so KAN is chosen as the fundamental classification network [38]. The specific network structure and parameter settings are in Table 3. During the model training process, the batch size is set to 16; the learning rate is set to 0.0001; the optimizer selects Adam; the epoch is set to 50 times; and the parameters of ISGDA are set according to the suggestions in reference [39].

Table 3. Network structure and parameters setting.

Number	Network layer	Type	Input Size	Output Size	Activation Function	Bias
1	Input	Reshape	(B, C, L)	(B, L×C)	-	-
2	Hidden Layer 1	Linear	L×C	512	Sigmoid	True
3	Hidden Layer 2	Linear	512	256	Sigmoid	True
4	Output	Linear	256	4	-	True

Note: Input denotes the input layer, Hidden denotes the hidden layer of the network, Output denotes the output layer, Type denotes the specific operation type used by each layer, Activation Function denotes the activation function used in the current layer, Bias denotes the bias term, B denotes the number of sample batches, C denotes the number of channels, and L denotes the signal length, and "-" denotes not applicable.

In the experiments discussed in this paper, the other detailed parameter settings of the proposed DACR model are shown in Table 4.

Table 4. DACR model detailed parameter settings.

Parameter Name	Notation	Parameter Value
Sampling frequency	fs	10 kHz
ISGDA signal-to-noise ratio	SNR	20 dB
Trajectory matrix window length	nfft	256
Correlation coefficient threshold of each component	-	0.8
Normalized mean square error decision threshold	-	0.001
Pseudo-label threshold	τ	0.95
Unlabeled loss weight	λ_u	1
Activation function	-	Sigmoid
Optimizer	-	Adam
Epochs	-	100
Batch size	-	16
Learning rate	lr	1×10^{-4}

(2) Analysis of results.

The effectiveness of DACR methodology is validated via contrasting it with other troubleshooting models with or without semi-supervised strategies using an experimental dataset of pump fault simulations. The core structures and detailed training configurations of different comparison models are shown in Table 5. First, the DACR approach is contrasted with three state-of-the-art semi-supervised learning approaches MixMatch [42], Pi-Model [43], and Mean Teacher [44], which are named MM-Kan, Pi-Kan, and MT-Kan, respectively, for ease of comparison. In MixMatch, it utilizes unlabeled data efficiently by mixing unlabeled data with labeled data through MixUp operation and imposing consistency regularization. In Pi-Model, it utilizes unlabeled data by generating similar outputs for different perturbations of the same input. In Mean teacher, for the same unlabeled input, separate predictions are made using the student model and the teacher model, and the difference between the two is calculated as a loss of consistency, thus providing more stable pseudo-labeling. These methods have been widely used in the field of image classification tasks under limited labeled samples. Also, considering ISGDA as an important component of DACR approach, it is compared with supervised learning models using only labeled data and only data augmentation, named LD-Kan, DA-Kan, respectively.

Table 5. Parameter settings for different compared model architectures.

Component	DACR	MM-Kan	Pi-Kan	MT-Kan	LD-Kan	DA-Kan
ISGDA Data Augmentation	Weak and Strong	Strong	Strong	Strong	None	Weak and Strong
Label post-processing	Pseudo-labeling	Sharpening	None	EMA	None	None
Consistency Loss	MSE (masked by threshold)	MSE with MixUp	MSE between input pair	MSE between student & teacher	None	None
Supervised Loss	CE	CE	CE	CE	CE	CE
Total Loss	CE + λ_u*CL	CE + λ_u*CL	CE + λ_u*CL	CE + λ_u*CL	CE	CE

In the research domain of fault diagnosis, the accuracy rate is usually considered as the basic statistical index to measure the effectiveness of diagnostic models [45]. In addition, this paper introduces three extra indicators, F1 score [46], precision, and recall [47,48], in order to assess the model properties in more depth, as shown below.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{27}$$

$$P = \frac{TP}{TP + FP} \tag{28}$$

$$R = \frac{TP}{FP + FN} \tag{29}$$

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{30}$$

where TP is the true case, TN is the true-negative case, FP is the false-positive case, and FN is the false-negative case. Acc is the accuracy ratio, P is the precision ratio, R is the recall ratio, and $F1$ is the F1 score. To decrease random error, ten experiments are performed for all contrasts.

According to the results in Figure 7a, it can be observed that the accuracy of DACR is always in the leading position in the ten trials. In particular, the maximum accuracy is reached in the fourth trial, which is about 3.8% above the minimum value in the fifth trial. In general, DACR shows less volatility, while the other five models show significant volatility. Figure 7b further demonstrates that the method consistently achieves the top precision ratio in all the trials, with a maximum peak precision ratio of 100% in the fourth trial. In contrast, the peak precision ratio of LD-Kan, DA-Kan, Pi-Kan, MM-Kan, and MT-Kan are lower than that of the method by 30.40%, 10.76%, 7.99%, 5.89%, and 10.88%, respectively. Meanwhile, Figure 7c demonstrates that the DACR method also ranks high in the recall metrics. The maximum recall for the fourth peak is 100%, while the range of volatility across experiments is small, with a relative difference of only 3.9% among the largest and smallest values. Lastly, the results of F1 scores are presented in Figure 7d, where DACR performs the best in all trials. The top F1 scores for the fourth trial are 100% and the mean standard deviation is only 0.38%, which indicates that the method possesses a superior and robust property.

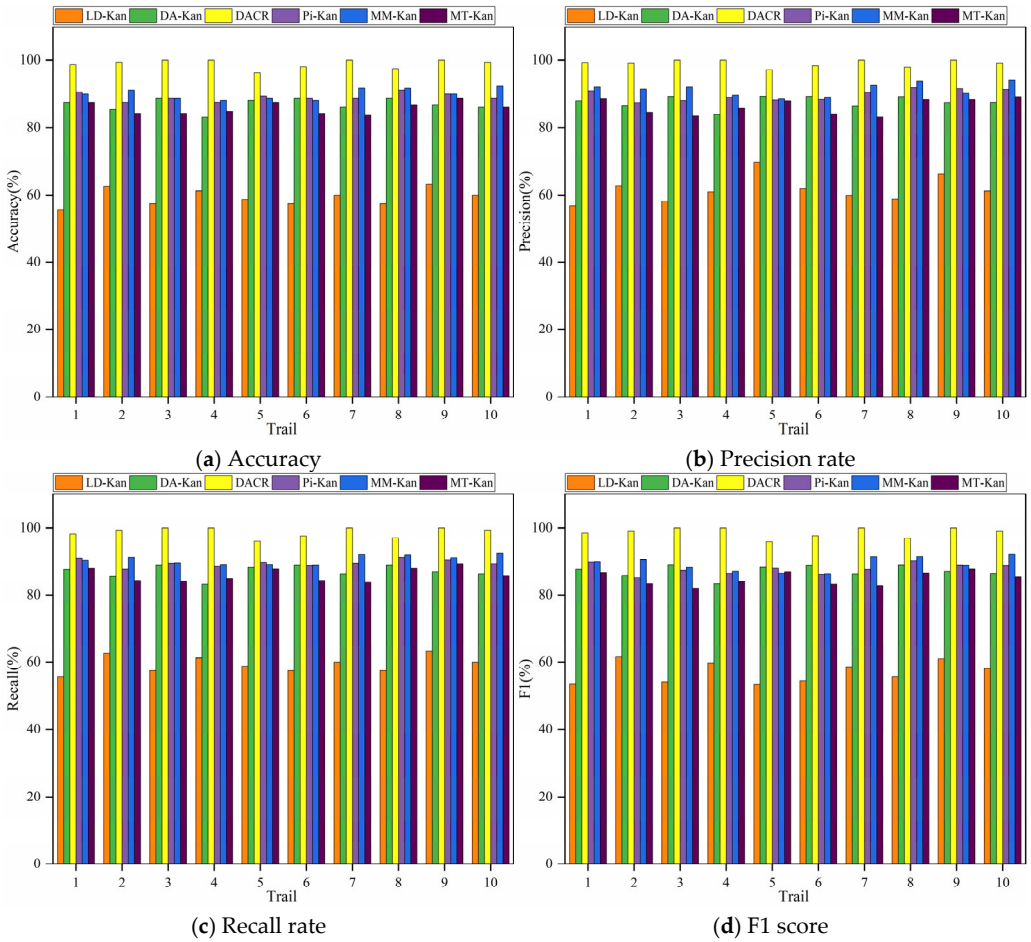


Figure 7. Four evaluation indicators for different methods in ten experiments.

Taken together, the DACR method performed well in all ten trials, and the four key indicators support this conclusion. The core of this approach's enhanced performance lies in the combination of a consistency strategy while applying additional perturbations to the temporal signal, which effectively solves the problem of scarcity of labeled samples and decreases the occurrence of model overfitting situations. The average and standard deviation of the four evaluation indicators in Figure 6 are displayed in Table 6. The mean accuracy of DACR is 98.94%, which is 39.56%, 11.94%, 9.81%, 8.81%, and 13.06% higher compared to LD-Kan, DA-Kan, Pi-Kan, MM-Kan, and MT-Kan, respectively. DACR equally outperforms the other five approaches with respect to precision, recall, and F1 scores. Notably, the approach has the lowest fluctuation in mean accuracy at 0.45%. These data demonstrate that the DACR method not only delivers superior performance but also retains excellent consistency.

Table 6. The mean of the four statistical indicators (%).

Metric	Methods					
	LD-Kan	**DA-Kan**	**Proposed**	**Pi-Kan**	**MM-Kan**	**MT-Kan**
Accuracy	59.38 ± 3.09	87.00 ± 0.88	98.94 ± 0.45	89.13 ± 1.32	90.13 ± 1.77	85.88 ± 0.88
Precision	61.65 ± 3.13	87.71 ± 0.32	99.13 ± 0.09	89.78 ± 0.33	91.40 ± 1.34	86.41 ± 0.34
Recall	59.38 ± 3.09	87.00 ± 0.88	98.79 ± 0.74	89.42 ± 1.33	90.49 ± 1.68	85.94 ± 1.47
F1	57.05 ± 3.17	87.07 ± 0.82	98.75 ± 0.38	87.74 ± 0.70	89.17 ± 1.71	84.81 ± 0.83

To demonstrate better visualization of the advantages of DACR methods in space distribution, the high dimensional characteristic spreads generated by various approaches are compared by T-SNE technique in Figure 8. In LD-Kan approach, the more obvious overlap among four classes of H, SPWF, SWF, and LBF, and the distribution of samples within the classes is more dispersed, which can easily cause classification mistakes. In the DA-Kan approach, the data in the SWF class is spread among the remaining three classes, and some of the data in the LBF class overlaps with the H and SWF classes. The characteristic distribution of the DACR approach is relatively better, although a few LBF samples are close to H class, but the majority of data maintains better within-class aggregation, and the border of different classes is clearer. In the Pi-Kan approach, the H class is heavily confounded with the LBF class and SWF class overlaps with other classes to varying degrees. The MM-Kan approach shows that a few samples in the LBF class are close to the H class, meanwhile there is a little conflation among the SPWF and SWF classes. The MT-Kan approach instead exhibits significant overlap between the H and LBF classes, as well as overlap between SWF and other classes. All these outcomes indicate that the DACR approach has a more centralized feature distribution and clearer boundaries among classes compared to other approaches.

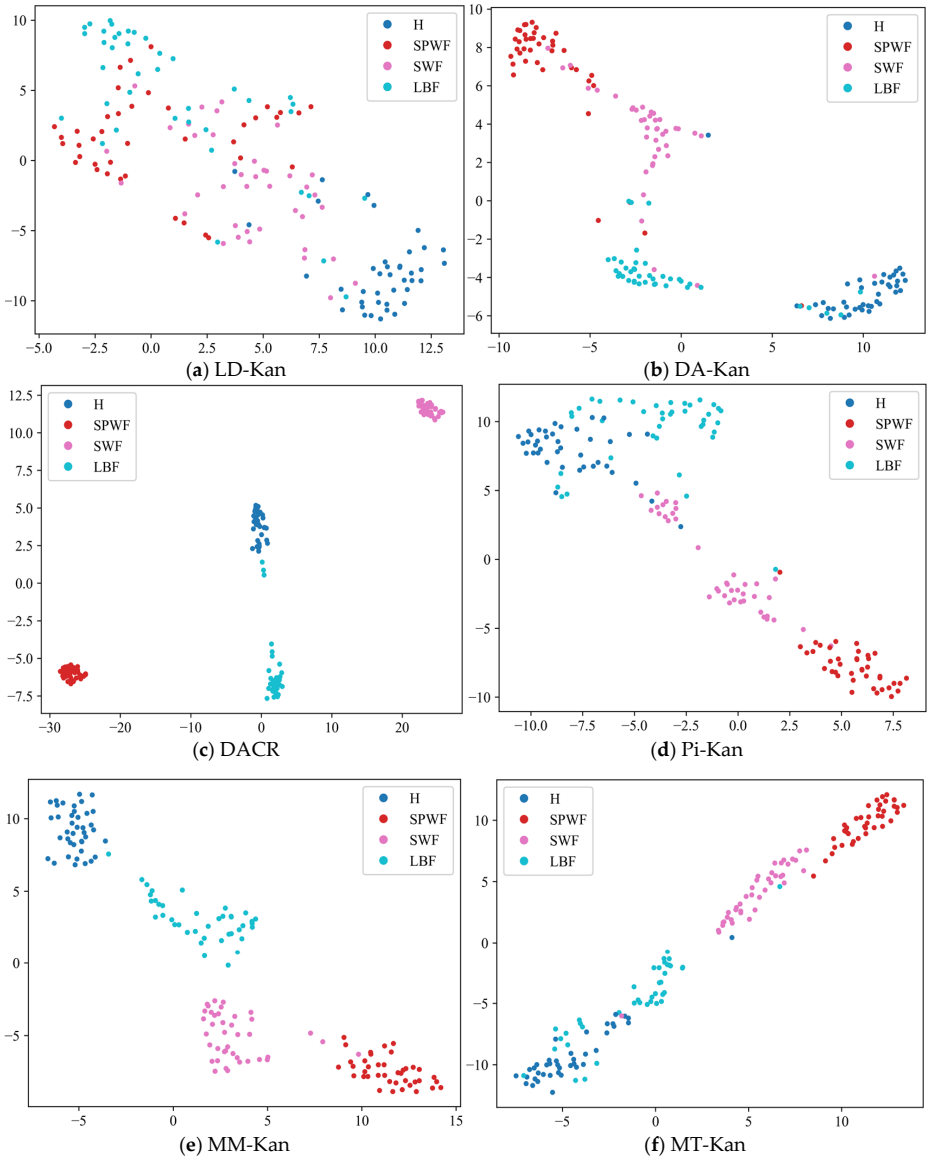
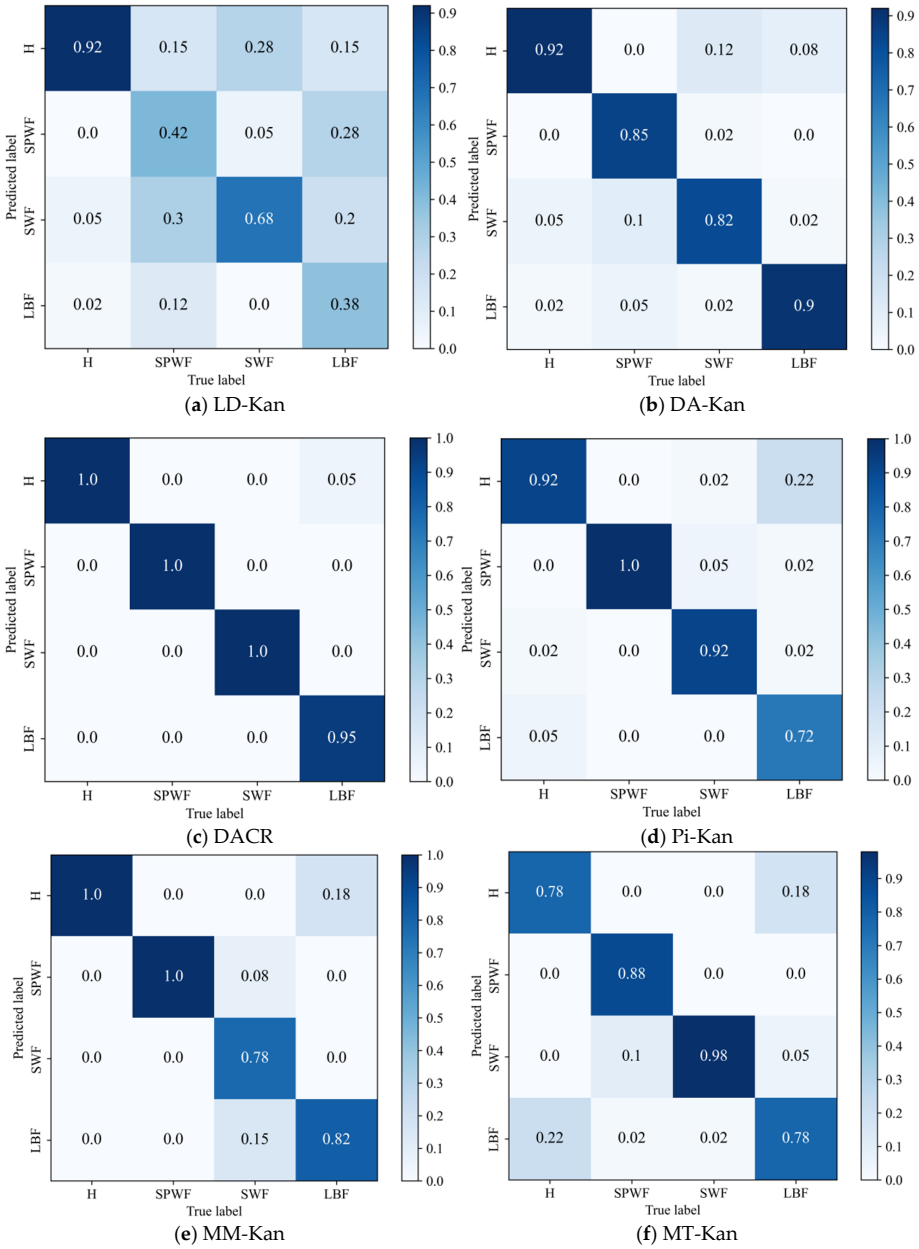


Figure 8. T-SNE features dimension reduction results of different approaches.

Figure 9 illustrates the confusion matrices for the LD-Kan, DA-Kan, Pi-Kan, MM-Kan, MT-Kan, and DACR approaches. For the LD-Kan approach, significant misclassification occurred for the SPWF, SWF, and LBF classes, mainly attributed to the insufficient number of labeled samples. The introduction of ISGDA as an additional perturbation in the DA-Kan approach provides limited enhancement, although it mitigates the misclassification due to inadequate labeled samples. Among the semi-supervised methods, Pi-Kan performs poorly on the H, SPWF, and LBF classes with significant misclassification. The MM-Kan approach is less effective in categorizing the SWF and LBF classes. The MT-Kan approach also encountered greater difficulties in identifying the H, SPWF and LBF classes. In comparison, the DACR approach has the best classification effect in all categories. It indicates that the designed ISGDA with semi-supervised consistency strategy can effectively extend the feature space of labeled data, while making full use of unlabeled data to optimize the model. As a result, the DACR approach realizes significant performance improvement and demonstrates superior fault diagnosis capability in the case of the scarcity of labeled samples.



(**a**) LD-Kan (**b**) DA-Kan
(**c**) DACR (**d**) Pi-Kan
(**e**) MM-Kan (**f**) MT-Kan

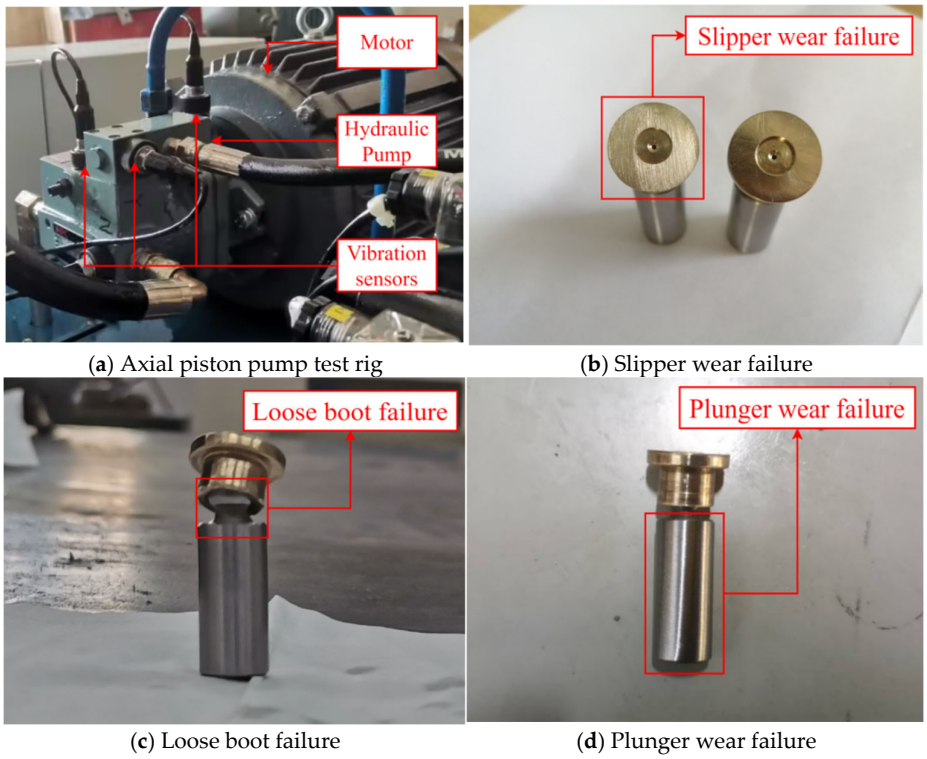
Figure 9. Confusion matrix for different approaches.

The performance of the proposed DACR approach is comprehensively evaluated by a range of property indicators, including accuracy, precision, recall, F1 value, T-SNE visualization, and confusion matrix. The experimental outcomes show that the DACR approach has a significant advantage in all the metrics. Further tests on Case 2 will follow to deeply analyze its effectiveness and stability under many different data conditions.

4.2. Case 2: Type P08-B3F-R-01 Fault Emulation Test Platform

(1) Overview of the test system and data.

The type P08-B3F-R-01 fault simulation test platform consists of axial piston pumps, acceleration sensors, AC motors, industrial control computer, and other components in Figure 10. Four hydraulic piston pump working situations are simulated separately: normal condition, sliding shoe abrasion, sliding shoe loosening, and plunger abrasion. The motor rotation frequency is set at a fixed speed of 1440 rpm, and the sampling frequency of the recorded data is 40 kHz during the test period. There are 249 examples per condition in Table 7, dividing the proportion of training samples and test samples into 8:2. To revert to the scarcity of labeled samples in engineering applications, the 199 training samples in each state have only 5%, that is 10 data labeled, and the rest are treated as unlabeled samples.



(**a**) Axial piston pump test rig (**b**) Slipper wear failure

(**c**) Loose boot failure (**d**) Plunger wear failure

Figure 10. Test platform and parts for different failure conditions.

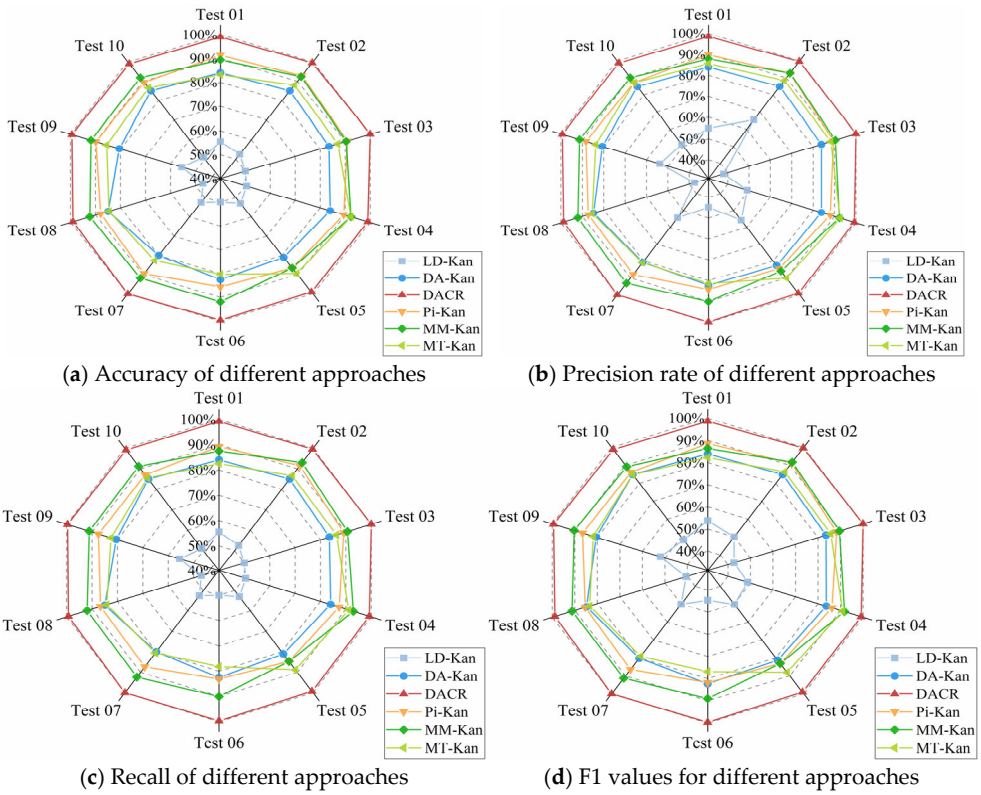
Table 7. Case 2: Statement of test data.

Label	State of Health	Explicit Explanation	Number of Training Datasets	Number of Test Datasets
0	H	State of health	199	50
1	SWF	Slipper wear failure	199	50
2	LBF	Loose boot failure	199	50
3	PWF	Plunger wear failure	199	50

(2) Analysis of test results.

The DACR approach is evaluated in this section using empirical data from different test platforms for a more comprehensive assessment of its advantages. Five methods, LD-Kan, DA-Kan, Pi-Kan, MM-Kan, and MT-Kan, are selected for contrast analysis. To minimize random mistakes, ten trials are conducted for all contrasts. For a detailed description of each approach, please consult Section 4.1.

Figure 11 illustrates the results of the four evaluation indicators of the different approaches over the ten trials, with the DACR approach performing the best in each test. The average and standard deviation of the evaluation indicators are summarized in Table 8. The mean accuracy rate of DACR is 99.37%, which is 47.62%, 16.32%, 10.05%, 8.31%, and 13.41% higher compared to LD-Kan, DA-Kan, Pi-Kan, MM-Kan, and MT-Kan, respectively. DACR equally outperforms the other five approaches with respect to precision, recall, and F1 scores. Notably, the approach has the lowest fluctuation in average accuracy at 0.01%. These data indicate that the DACR approach is not only capable of delivering superior performance but also excels in consistency and stability.



(**a**) Accuracy of different approaches (**b**) Precision rate of different approaches

(**c**) Recall of different approaches (**d**) F1 values for different approaches

Figure 11. Four evaluation indicators for different approaches in ten experiments.

Table 8. Mean results for the four statistical indicators (%).

Metric	Methods					
	LD-Kan	**DA-Kan**	**Proposed**	**Pi-Kan**	**MM-Kan**	**MT-Kan**
Accuracy	51.75 ± 3.18	83.05 ± 0.71	99.37 ± 0.01	89.32 ± 1.36	91.06 ± 1.70	85.96 ± 2.38
Precision	50.48 ± 2.71	83.21 ± 0.71	99.31 ± 0.14	87.92 ± 1.81	90.57 ± 1.63	86.36 ± 1.10
Recall	51.75 ± 3.18	83.05 ± 0.71	99.52 ± 0.02	87.36 ± 1.56	90.54 ± 2.63	84.63 ± 2.17
F1	48.10 ± 3.79	82.98 ± 0.72	99.32 ± 0.12	86.35 ± 1.70	89.34 ± 2.02	83.91 ± 1.82

To demonstrate better visualization of the advantages of the DACR approach in spatial distribution, the high dimensional characteristic spreads generated by various approaches

are compared by T-SNE technique. Figure 12 illustrates the feature dimension decrease results of LD-Kan, DA-Kan, Pi-Kan, MM-Kan, MT-Kan, and DACR approaches. The features of the DACR approach display a more obvious clustering effect on the spatial distribution and have a clearer category differentiation compared to the other approaches.

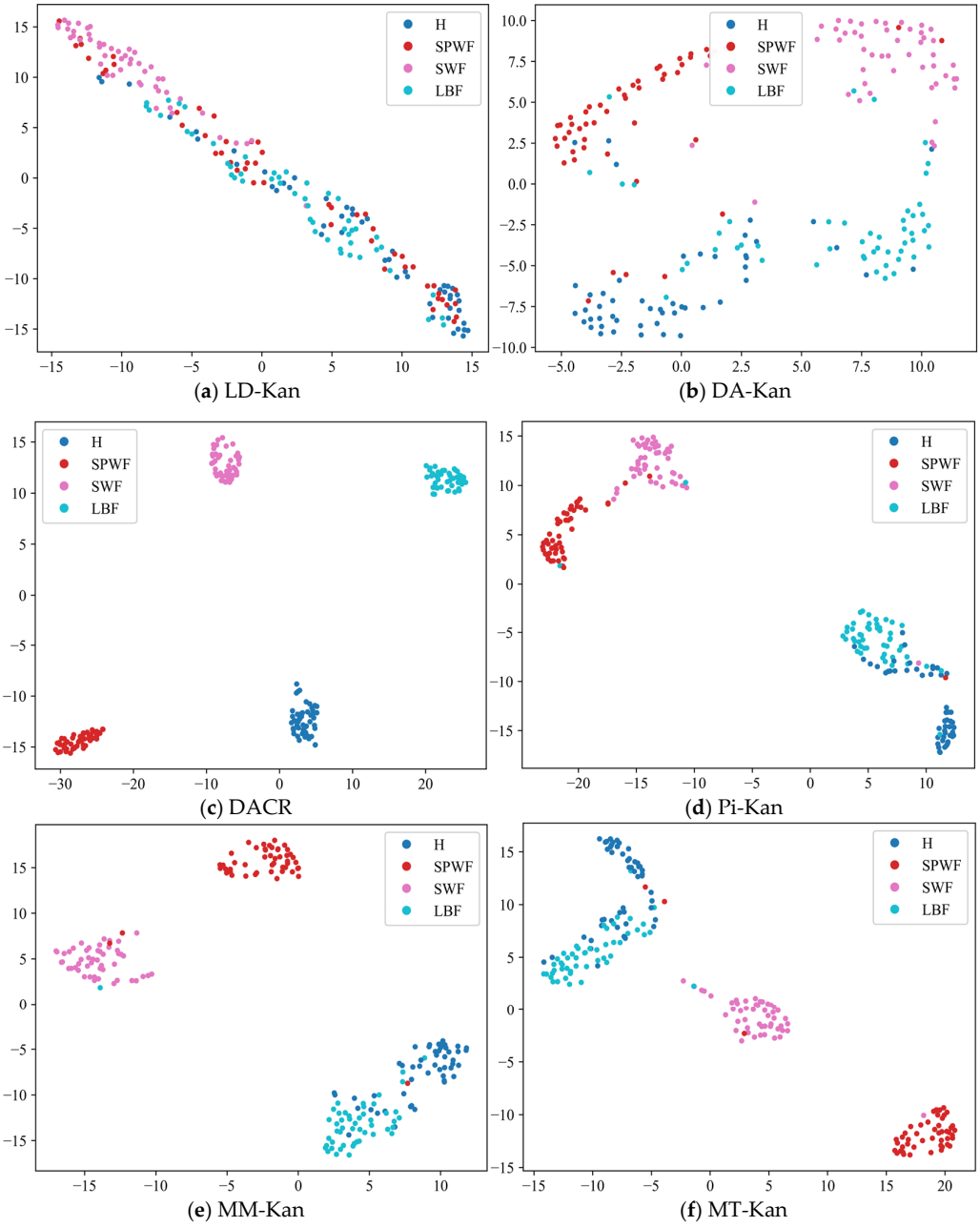
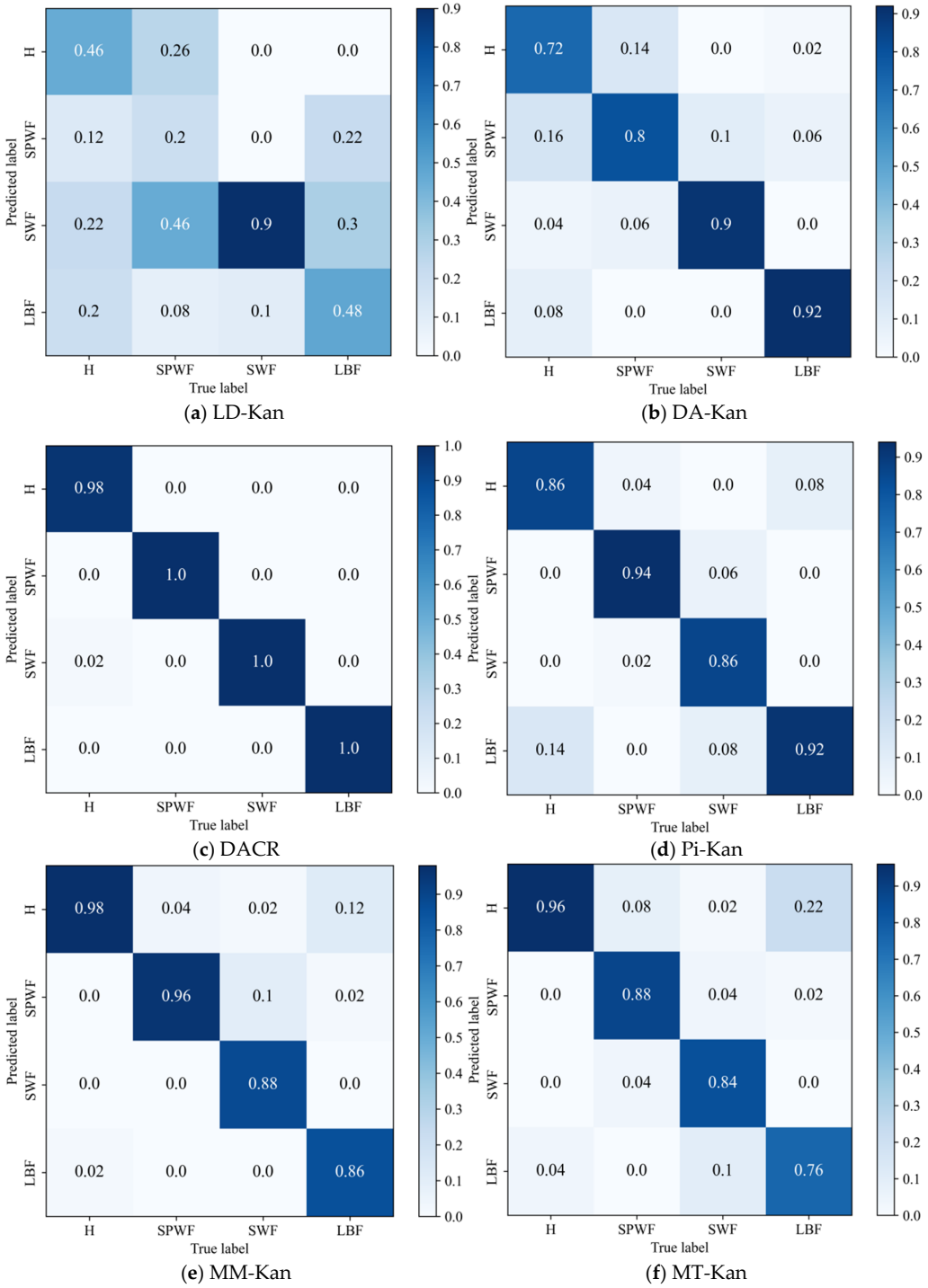


Figure 12. Visualization of feature dimension reduction by different methods.

Figure 13 illustrates the confusion matrices of the LD-Kan, DA-Kan, Pi-Kan, MM-Kan, MT-Kan, and DACR approaches to visually compare the diagnostic effectiveness of each model. The results indicate that the DACR approach significantly outperforms the other models in terms of classification accuracy in different categories, displaying excellent diagnostic capability.



(**a**) LD-Kan (**b**) DA-Kan

(**c**) DACR (**d**) Pi-Kan

(**e**) MM-Kan (**f**) MT-Kan

Figure 13. Confusion matrix of different approaches.

To verify the practical applicability of the proposed DACR model, this paper analyzes its computational complexity. For the two experimental cases, four average quantitative indicators are recorded: the number of floating-point operations, the total number of model parameters, memory usage, and test time. As shown in Table 9, the FLOPs are 0.25 G, the memory usage is 96.51 MB, and the test time is only 0.56 s. The above results demonstrate that the DACR method has high computational efficiency and meets engineering requirements such as online condition monitoring of hydraulic pumps.

Table 9. Average computational complexity of the DACR model in two cases.

Metric	FLOPs (G)	Params (M)	Memory (MB)	Testing Time (s)
Value	0.25	25.30	96.51	0.56

4.3. DACR Model Performance with Distinct Labeled Sample Proportions

This session focuses on analyzing the performance variation in the DACR approach with different proportions of labeled data. The same dataset and various parameter settings as in Case 1 are used. The trial is designed to cover five labeling ratios of 1%, 2%, 5%, 10%, and 20%, and ten replications of the trial are conducted for each ratio. Figure 14 illustrates that DACR approach continues to rise in the mean value of each statistical metric as the proportion of labels rises. The data in Table 10 further demonstrate that when the proportion of labeled samples reaches 5%, 10%, and 20%, all the evaluation indexes of the DACR approach are more than 95%, showing strong stability. Notably, despite the labeling ratio of only 1% in the extreme case, its mean accuracy still reaches 74.88%. And the accuracy is enhanced by 14.19% when the labeling proportion is raised to 2%. In order to verify the performance of DACR approach more comprehensively, its robustness performance under different noise levels will be explored subsequently.

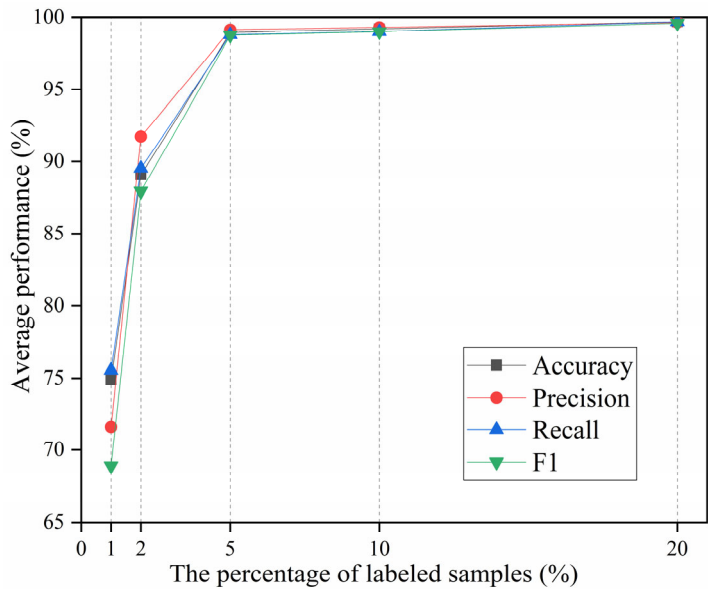


Figure 14. Trends with different proportions of labeled samples.

Table 10. Mean outcomes for different labeling sample proportions (%).

Metric	The Percentage of Labeled Samples				
	1%	2%	5%	10%	20%
Accuracy	74.88 ± 1.77	89.07 ± 3.09	98.94 ± 0.45	99.19 ± 1.32	99.69 ± 0.44
Precision	71.59 ± 6.60	91.75 ± 3.54	99.13 ± 0.09	99.30 ± 1.10	99.63 ± 0.59
Recall	75.55 ± 1.82	89.52 ± 3.15	98.79 ± 0.74	99.00 ± 1.62	99.69 ± 0.44
F1	68.91 ± 0.01	87.94 ± 3.99	98.75 ± 0.38	98.99 ± 1.58	99.58 ± 0.61

4.4. DACR Model Performance Under Different Noise Levels

This session analyzes the performance variation in the DACR approach under different noise levels, using the same dataset and various parameter settings as in Case 1. The experimental is designed to cover five different strengths of signal-to-noise ratios, namely

−10 dB, −5 dB, 0 dB, 5 dB, and 10 dB. The mean result variation trend of ten trials for each evaluation index is illustrated in Figure 15.

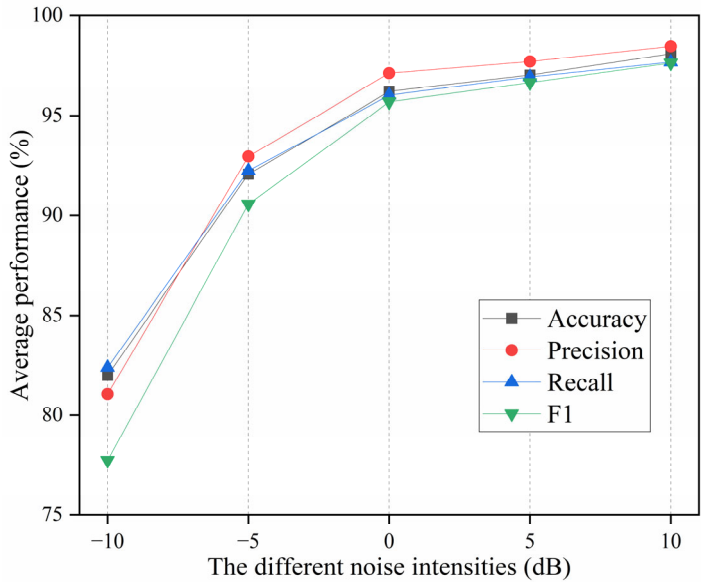


Figure 15. Trend of averaged experimental results for different signal-to-noise ratios.

Figure 15 illustrates that as the signal-to-noise ratio varies, the mean value of the DACR approach continues to rise for each statistical metric. Based on the specific data in Table 11, the DACR approach achieves more than 90% in all four performance indicators when the signal-to-noise ratio exceeds −5 dB, and the volatility is less than 4%. Moreover, the F1 values for the rest of the different SNR conditions improved by 12.83%, 17.95%, 18.89%, and 19.87%, respectively, compared to those at −10 dB SNR. These outcomes demonstrate that DACR approach maintains strong diagnostic capabilities in the face of noise disturbances of different intensities, which further validates its excellent robustness.

Table 11. Experimental results at different signal-to-noise ratios (%).

Metric	Different Noise Intensities (dB)				
	−10	**−5**	**0**	**5**	**10**
Accuracy	82.00 ± 5.30	92.07 ± 2.21	96.19 ± 0.45	97.00 ± 0.88	98.07 ± 0
Precision	81.06 ± 8.67	92.96 ± 3.44	97.09 ± 0.25	97.68 ± 0.47	98.44 ± 0.01
Recall	82.40 ± 5.45	92.25 ± 2.06	96.00 ± 0.45	96.90 ± 0.88	97.65 ± 0
F1	77.73 ± 7.91	90.56 ± 3.24	95.68 ± 0.39	96.62 ± 1.21	97.60 ± 0

5. Conclusions

Aiming to address the problem of the scarcity of labeled samples in hydraulic pump troubleshooting, this paper innovatively presents a semi-supervised learning approach based on DACR, which effectively utilizes the unlabeled samples and prevents the overfitting phenomenon in the process of model training. The validity and applicability of the proposed approach is verified by performing tests on two types of different pump datasets. The specific conclusions are summarized as below:

(1) The results of the comparison trials with other approaches indicate that the DACR approach proposed in this research has excellent classification capability for networks trained on pump class datasets under limited labeled sample conditions. In ten trials, the DACR approach is ahead of other approaches in accuracy, precision, recall, and F1 value performance, while the overall volatility is kept at the lowest level.

(2) The results from the trial analysis of the model performance under different label proportions and different signal-to-noise ratios reveal that the DACR approach is capable of maintaining high diagnostic performance while possessing good robustness under low label sample proportions.
(3) In terms of technology diffusion, the DACR approach is not only suitable for fault diagnosis tasks under limited labeling samples in dealing with other rotating mechanical devices, but also able to be integrated with various classification model structures according to the actual application requirements, demonstrating a promising application prospect.

Author Contributions: Conceptualization, Z.Z. and C.A.; methodology, J.Y., S.L., Z.Z. and Y.Z.; investigation, Z.Z. and Y.Z.; validation, J.Y., Z.Z. and Y.Z.; resources, C.A. and W.J.; writing—original draft preparation, J.Y.; writing—review and editing, J.Y. and S.L.; supervision, S.L.; project administration, S.L.; funding acquisition, S.L., C.A. and W.J. All authors have read and agreed to the published version of the manuscript.

Funding: The work was funded by the National Natural Science Foundation of China (Nos. 52275069 and 52275067), the S&T Program of Hebei (Grant No. 236Z4502G), and the Bureau of Science and Technology of Hebei Province, China, grant number (No. E2021203020).

Data Availability Statement: Data will be made available on request.

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

References

1. Xu, X.; Zhang, J.; Huang, W.; Yu, B.; Lyu, F.; Zhang, X.; Xu, B. The loose slipper fault diagnosis of variable-displacement pumps under time-varying operating conditions. *Reliab. Eng. Syst. Saf.* **2024**, *252*, 110448. [CrossRef]
2. Guo, J.; Liu, Y.; Yang, R.; Sun, W.; Xiang, J. A simulation-driven difference mode decomposition method for fault diagnosis in axial piston pumps. *Adv. Eng. Inform.* **2024**, *62*, 102624. [CrossRef]
3. Xu, Z.; Wang, Z.; Gao, C.; Zhang, K.; Lv, J.; Wang, J.; Liu, L. A digital twin system for centrifugal pump fault diagnosis driven by transfer learning based on graph convolutional neural networks. *Comput. Ind.* **2024**, *163*, 104155. [CrossRef]
4. Prasshanth, C.V.; Venkatesh, S.N.; Mahanta, T.K.; Sakthivel, N.R.; Sugumaran, V. Fault diagnosis of monoblock centrifugal pumps using pre-trained deep learning models and scalogram images. *Eng. Appl. Artif. Intell.* **2024**, *136*, 109022. [CrossRef]
5. Li, Z.; Liu, Z.; Zuo, M. Homotypic multi-source mixed signal decomposition based on maximum time-shift kurtosis for drilling pump fault diagnosis. *Mech. Syst. Signal Process.* **2024**, *221*, 111724. [CrossRef]
6. Varejão, F.M.; Mello, L.H.S.; Ribeiro, M.P.; Oliveira-Santos, T.; Rodrigues, A.L. An open source experimental framework and public dataset for vibration-based fault diagnosis of electrical submersible pumps used on offshore oil exploration. *Knowl.-Based Syst.* **2024**, *288*, 111452. [CrossRef]
7. Fu, S.; Zou, L.; Wang, Y.; Lin, L.; Lu, Y.; Zhao, M.; Guo, F.; Zhong, S. DCSIAN: A novel deep cross-scale interactive attention network for fault diagnosis of aviation hydraulic pumps and generalizable applications. *Reliab. Eng. Syst. Saf.* **2024**, *249*, 110246. [CrossRef]
8. Li, Y.; Zhang, L.; Liang, P.; Wang, X.; Wang, B.; Xu, L. Semi-supervised meta-path space extended graph convolution network for intelligent fault diagnosis of rotating machinery under time-varying speeds. *Reliab. Eng. Syst. Saf.* **2024**, *251*, 110363. [CrossRef]
9. Zhong, Q.; Xu, E.; Shi, Y.; Jia, T.; Ren, Y.; Yang, H.; Li, Y. Fault diagnosis of the hydraulic valve using a novel semi-supervised learning method based on multi-sensor information fusion. *Mech. Syst. Signal Process.* **2023**, *189*, 110093. [CrossRef]
10. Xu, H.; Wang, X.; Huang, J.; Zhang, F.; Chu, F. Semi-supervised multi-sensor information fusion tailored graph embedded low-rank tensor learning machine under extremely low labeled rate. *Inf. Fusion* **2024**, *105*, 102222. [CrossRef]
11. Huang, Z.; Li, K.; Xu, Z.; Yin, R.; Yang, Z.; Mei, W.; Bing, S. STP-Model: A semi-supervised framework with self-supervised learning capabilities for downhole fault diagnosis in sucker rod pumping systems. *Eng. Appl. Artif. Intell.* **2024**, *135*, 108802. [CrossRef]
12. Fu, X.; Tao, J.; Jiao, K.; Liu, C. A novel semi-supervised prototype network with two-stream wavelet scattering convolutional encoder for TBM main bearing few-shot fault diagnosis. *Knowl.-Based Syst.* **2024**, *286*, 111408. [CrossRef]
13. Liang, P.; Xu, L.; Shuai, H.; Yuan, X.; Wang, B.; Zhang, L. Semi-supervised subdomain adaptation graph convolutional network for fault transfer diagnosis of rotating machinery under time-varying speeds. *IEEE/ASME Trans. Mechatron.* **2024**, *29*, 730–741. [CrossRef]

14. Yao, X.; Lu, X.; Jiang, Q.; Shen, Y.; Xu, F.; Zhu, Q. SSPENet: Semi-supervised prototype enhancement network for rolling bearing fault diagnosis under limited labeled samples. *Adv. Eng. Inform.* **2024**, *61*, 102560. [CrossRef]
15. Han, T.; Xie, W.; Pei, Z. Semi-supervised adversarial discriminative learning approach for intelligent fault diagnosis of wind turbine. *Inf. Sci.* **2023**, *648*, 119496. [CrossRef]
16. Deng, C.; Deng, Z.; Miao, J. Semi-supervised ensemble fault diagnosis method based on adversarial decoupled auto-encoder with extremely limited labels. *Reliab. Eng. Syst. Saf.* **2024**, *242*, 109740. [CrossRef]
17. Yan, S.; Shao, H.; Xiao, Y.; Zhou, J.; Xu, Y.; Wan, J. Semi-supervised fault diagnosis of machinery using LPS-DGAT under speed fluctuation and extremely low labeled rates. *Adv. Eng. Inform.* **2022**, *53*, 101648. [CrossRef]
18. Zhang, T.; Li, C.; Chen, J.; He, S.; Zhou, Z. Feature-level consistency regularized Semi-supervised scheme with data augmentation for intelligent fault diagnosis under small samples. *Mech. Syst. Signal Process.* **2023**, *203*, 110747. [CrossRef]
19. Ramírez-Sanz, J.M.; Maestro-Prieto, J.A.; Arnaiz-González, Á.; Bustillo, A. Semi-supervised learning for industrial fault detection and diagnosis: A systemic review. *ISA Trans.* **2023**, *143*, 255–270. [CrossRef]
20. Zhang, L.; Wang, B.; Liang, P.; Yuan, X.; Li, N. Semi-supervised fault diagnosis of gearbox based on feature pre-extraction mechanism and improved generative adversarial networks under limited labeled samples and noise environment. *Adv. Eng. Inform.* **2023**, *58*, 102211. [CrossRef]
21. Miao, J.; Deng, Z.; Deng, C.; Chen, C. Boosting efficient attention assisted cyclic adversarial auto-encoder for rotating component fault diagnosis under low label rates. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108499. [CrossRef]
22. He, Y.; He, D.; Lao, Z.; Jin, Z.; Miao, J.; Lai, Z.; Chen, Y. Few-shot fault diagnosis of turnout switch machine based on flexible semi-supervised meta-learning network. *Knowl.-Based Syst.* **2024**, *294*, 111746. [CrossRef]
23. Ozdemir, R.; Koc, M. On the enhancement of semi-supervised deep learning-based railway defect detection using pseudo-labels. *Expert Syst. Appl.* **2024**, *251*, 124105. [CrossRef]
24. Azar, K.; Hajiakhondi-Meybodi, Z.; Naderkhani, F. Semi-supervised clustering-based method for fault diagnosis and prognosis: A case study. *Reliab. Eng. Syst. Saf.* **2022**, *222*, 108405. [CrossRef]
25. Su, Z.; Zhang, J.; Xu, H.; Zou, J.; Fan, S. Deep semi-supervised transfer learning method on few source data with sensitivity-aware decision boundary adaptation for intelligent fault diagnosis. *Expert Syst. Appl.* **2024**, *249*, 123714. [CrossRef]
26. Lu, F.; Tong, Q.; Jiang, X.; Feng, Z.; Xu, J.; Wang, X.; Huo, J. A deep targeted transfer network with clustering pseudo-label learning for fault diagnosis across different Machines. *Mech. Syst. Signal Process.* **2024**, *213*, 111344. [CrossRef]
27. Kumar, D.D.; Fang, C.; Zheng, Y.; Gao, Y. Semi-supervised transfer learning-based automatic weld defect detection and visual inspection. *Eng. Struct.* **2023**, *292*, 116580. [CrossRef]
28. Yu, T.; Li, C.; Huang, J.; Xiao, X.; Zhang, X.; Li, Y.; Fu, B. ReF-DDPM: A novel DDPM-based data augmentation method for imbalanced rolling bearing fault diagnosis. *Reliab. Eng. Syst. Saf.* **2024**, *251*, 110343. [CrossRef]
29. Kulevome, D.K.B.; Wang, H.; Cobbinah, B.M.; Mawuli, E.S.; Kumar, R. Effective time-series Data Augmentation with Analytic Wavelets for bearing fault diagnosis. *Expert Syst. Appl.* **2024**, *249*, 123536. [CrossRef]
30. Tian, J.; Jiang, Y.; Zhang, J.; Luo, H.; Yin, S. A novel data augmentation approach to fault diagnosis with class-imbalance problem. *Reliab. Eng. Syst. Saf.* **2024**, *243*, 109832. [CrossRef]
31. Mueller, P.N. Attention-enhanced conditional-diffusion-based data synthesis for data augmentation in machine fault diagnosis. *Eng. Appl. Artif. Intell.* **2024**, *131*, 107696. [CrossRef]
32. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
33. Martins, D.H.; de Lima, A.A.; Gutiérrez, R.H.; Pestana-Viana, D.; Netto, S.L.; Vaz, L.A.; da Silva, E.A.; Haddad, D.B. Improved variational mode decomposition for combined imbalance-and-misalignment fault recognition and severity quantification. *Eng. Appl. Artif. Intell.* **2023**, *124*, 106516. [CrossRef]
34. Wang, L.; Liu, Z. An improved local characteristic-scale decomposition to restrict end effects, mode mixing and its application to extract incipient bearing fault signal. *Mech. Syst. Signal Process.* **2021**, *156*, 107657. [CrossRef]
35. Ma, Y.; Cheng, J.; Wang, P.; Wang, J.; Yang, Y. A novel Lanczos quaternion singular spectrum analysis method and its application to bevel gear fault diagnosis with multi-channel signals. *Mech. Syst. Signal Process.* **2022**, *168*, 108679. [CrossRef]
36. Wang, N.; Ma, P.; Wang, X.; Wang, C.; Zhang, H. Detection of unknown bearing faults using re-weighted symplectic geometric node network characteristics and structure analysis. *Expert Syst. Appl.* **2023**, *215*, 119304. [CrossRef]
37. Yu, B.; Cao, N.; Zhang, T. A novel signature extracting approach for inductive oil debris sensors based on symplectic geometry mode decomposition. *Measurement* **2021**, *185*, 110056. [CrossRef]
38. Liu, Z.; Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Soljačić, M.; Hou, T.Y.; Tegmark, M. Kan: Kolmogorov-arnold networks. *arXiv* **2024**, arXiv:2404.19756.
39. Pan, H.; Yang, Y.; Li, X.; Zheng, J.; Cheng, J. Symplectic geometry mode decomposition and its application to rotating machinery compound fault diagnosis. *Mech. Syst. Signal Process.* **2019**, *114*, 189–211. [CrossRef]
40. Wang, S.; Hu, J.; Du, Y.; Yuan, X.; Xie, Z.; Liang, P. WCFormer: An interpretable deep learning framework for heart sound signal analysis and automated diagnosis of cardiovascular diseases. *Expert Syst. Appl.* **2025**, *276*, 127238. [CrossRef]

41. Xu, J.; Qu, J. Capacity estimation of lithium-ion battery based on soft dynamic time warping, stratified random sampling and pruned residual neural networks. *Eng. Appl. Artif. Intell.* **2024**, *138*, 109278. [CrossRef]
42. Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; Raffel, C. Mixmatch: A holistic approach to semi-supervised learning. *arXiv* **2019**, arXiv:1905.02249.
43. Laine, S.; Aila, T. Temporal ensembling for semi-supervised learning. *arXiv* **2016**, arXiv:1610.02242.
44. Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *arXiv* **2017**, arXiv:1703.01780.
45. Fu, S.; Lin, L.; Wang, Y.; Zhao, M.; Guo, F.; Zhong, S.; Liu, Y. High imbalance fault diagnosis of aviation hydraulic pump based on data augmentation via local wavelet similarity fusion. *Mech. Syst. Signal Process.* **2024**, *209*, 111115. [CrossRef]
46. Tang, S.; Zhu, Y.; Yuan, S. A novel adaptive convolutional neural network for fault diagnosis of hydraulic piston pump with acoustic images. *Adv. Eng. Inform.* **2022**, *52*, 101554. [CrossRef]
47. Qiu, Z.; Li, W.; Tang, T.; Wang, D.; Wang, Q. Denoising graph neural network based hydraulic component fault diagnosis method. *Mech. Syst. Signal Process.* **2023**, *204*, 110828. [CrossRef]
48. Huang, X.; Zhang, J.; Huang, W.; Lyu, F.; Xu, H.; Xu, B. Multi-output sparse Gaussian process based fault detection for a variable displacement pump under random time-variant working conditions. *Mech. Syst. Signal Process.* **2024**, *211*, 111191. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.