energies

MDPI

Article

Open-Set Bearing Fault Diagnosis Under Variable Operating Conditions Using Mahalanobis-Weighted Adversarial Adaptation

Xin-Ming Liu [1,2], Hao Chen [1,*], Rui-Ming Zhang [3] and Shi-Ji Fang [1]

[1] Faculty of Electrical and Control Engineering, Liaoning Technical University, Huludao 125105, China; liuxin_ming@163.com (X.-M.L.); 18841814172@163.com (S.-J.F.)
[2] Ordos Research Institute, Liaoning Technical University, Ordos 017000, China
[3] Shijiazhuang Luancheng District Power Supply Branch, State Grid Hebei Electric Power Company, Shijiazhuang 051430, China; 18231039027@163.com
* Correspondence: 17639641290@163.com

Abstract

Rolling bearings are key components in electric motor-driven energy equipment, such as motors, pumps, compressors, fans, and generator systems. Their unexpected failures may reduce energy conversion efficiency, increase maintenance costs, and threaten the operational reliability of energy systems. However, in practical condition monitoring, electric machinery usually operates under variable working conditions, and unknown fault modes that are absent from the source domain may appear in the target domain, which leads to negative transfer and unreliable diagnostic decisions for conventional unsupervised domain adaptation methods. To address this problem, this paper proposes an open-set bearing fault diagnosis method based on a Weighted Adversarial Adaptive Network (WAAN). Specifically, a Mahalanobis distance-based sample weighting strategy is introduced to measure the deviation between target-domain samples and source-domain known-class distributions, thereby suppressing the forced alignment of potential unknown samples during adversarial adaptation. In addition, a sliding-window quantile-based dynamic threshold mechanism is developed to adaptively identify unknown fault categories without relying on manually fixed thresholds. Experiments on the CWRU bearing dataset and a laboratory motor-bearing platform demonstrate that WAAN achieves a known-class classification accuracy of 99.72% and an unknown-class detection rate of 76.25% across six open-set transfer tasks. These results indicate the potential of the proposed method for open-set bearing fault diagnosis under the investigated operating conditions, while its generalization to other machinery and more challenging environments requires further validation.

Keywords: fault diagnosis; open-set domain adaptation; domain adaptation; condition monitoring; Mahalanobis distance; intelligent fault diagnosis

Check for updates

Academic Editor: Marcin Kaminski

Received: 25 September 2026
Revised: 12 October 2026
Accepted: 15 October 2026
Published: 22 October 2026

Copyright: © 2026 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.

1. Introduction

In recent years, fault diagnosis methods based on unsupervised domain adaptation (UDA) have achieved significant progress in cross-domain transfer learning [1,2]. Most existing methods assume that the source and target domains share the same label space, meaning that the training and testing data cover identical machine health states. This assumption enables knowledge transfer by minimizing the distribution discrepancy between the source and target domains. However, this assumption is often violated in practical industrial scenarios. Due to limitations in data acquisition cost and labor, it is difficult to

obtain sufficient labeled data covering all possible health states, resulting in limited source-domain category coverage. During the testing phase, new fault modes may emerge, leading to a more challenging and realistic Open-Set Domain Adaptation (OSDA) problem [3,4], where the label space of the source domain is a subset of that of the target domain.

Traditional domain adaptation methods typically use kernel functions and other techniques to map features from the source and target domains to the same feature space and align their distributions to enhance inter-domain feature similarity and assist the model in transfer diagnosis. In recent years, deep learning-based methods have been widely applied in bearing fault diagnosis, demonstrating superior performance in feature extraction and fault classification compared with traditional approaches. Comprehensive reviews have summarized recent advancements in deep learning-based bearing fault diagnosis, highlighting the rapid development of intelligent diagnostic models and their challenges in practical applications [5]. Recent studies have explored advanced architectures such as CNN-based feature extraction combined with generative models to enhance diagnostic performance under complex working conditions and improve robustness against noise and data imbalance [6]. Attention mechanisms and hybrid deep learning frameworks have also been introduced to capture both temporal and spatial features from nonstationary signals, enabling more accurate representation of fault characteristics [7]. In addition, generative adversarial networks and improved residual networks have been employed to address challenges such as small sample size, noise interference, and data distribution discrepancies, thereby improving model generalization ability [8]. Furthermore, novel network structures such as KAN-enhanced CNN models have been proposed to improve feature representation capability and learning efficiency in limited data scenarios [9].

Despite these advances, open-set domain adaptation remains challenging when unknown fault categories are present in the target domain. Existing studies have explored various strategies to mitigate negative transfer and improve unknown-class detection. For example, instance-level weighted adversarial learning [3] and multiple-metric weighting strategies [10] have been investigated to improve selective knowledge transfer. Furthermore, knowledge-weighted adversarial adaptation and class alignment learning [11] have been explored to enhance feature alignment across different domains. In addition, soft unknown-class rejection [12] and adversarial learning-based open-set bearing fault diagnosis [13] have been studied to address the influence of unknown categories.

Although these approaches have advanced open-set domain adaptation, accurately estimating the transferability of target samples and establishing reliable unknown-class rejection criteria under changing operating conditions remain challenging. In particular, prediction-confidence-based criteria may be affected by overconfident predictions for unknown samples, while inappropriate adversarial alignment may lead to negative transfer. Therefore, this study focuses on two related challenges: (1) reducing the alignment contributions of potentially unknown target samples by considering their distributional deviations from known source-domain classes; and (2) establishing an adaptive unknown-class rejection criterion that reduces dependence on a manually fixed confidence threshold.

To address these challenges, this paper proposes a Weighted Adversarial Adaptive Network (WAAN) for open-set bearing fault diagnosis under variable operating conditions. The proposed method integrates Mahalanobis-distance-based sample weighting with a sliding-window quantile-based dynamic threshold mechanism. Specifically, Mahalanobis distance is employed to measure the deviation of target-domain features from known source-domain class distributions while accounting for feature covariance. The resulting sample weights regulate the contributions of target samples to adversarial feature alignment, thereby reducing the influence of potentially unknown faults. Meanwhile, the sliding-window quantile-based threshold adaptively adjusts the unknown-class rejection

criterion according to recent prediction-confidence statistics. Through the integration of these two mechanisms, WAAN aims to improve known-class transferability while mitigating negative transfer and enhancing unknown-fault detection.

The principal contribution of WAAN lies in the coordinated integration of distribution-aware sample weighting and adaptive unknown-class rejection within an open-set domain adaptation framework. Specifically, Mahalanobis-distance-based weighting regulates target-sample contributions during adversarial feature alignment, while sliding-window quantile thresholding governs unknown-class rejection during inference. These mechanisms address complementary stages of the diagnostic process, rather than constituting entirely new standalone distance or threshold estimation techniques.

The main contributions of this study are summarized as follows:

(1) A unified open-set domain adaptation framework is developed by integrating Mahalanobis-distance-based sample weighting during adversarial feature alignment with adaptive unknown-class rejection during inference, thereby jointly addressing inappropriate feature alignment and unknown-fault detection under variable operating conditions.

(2) A Mahalanobis-distance-based sample weighting mechanism is incorporated into adversarial domain adaptation. By considering the covariance structure of known source-domain feature distributions, the proposed mechanism estimates the distributional deviations of target samples and assigns lower adversarial weights to samples that are less likely to belong to known categories, thereby reducing inappropriate cross-domain alignment.

(3) A sliding-window quantile-based dynamic threshold strategy is integrated into the open-set diagnostic decision process. Unlike a manually fixed confidence threshold, the proposed strategy updates the unknown-class rejection criterion according to recent prediction-confidence statistics, improving the adaptability of diagnostic decisions under variable operating conditions.

The primary objective of this study is to achieve reliable open-set bearing fault diagnosis under variable operating conditions by preserving known-class classification performance while improving unknown-fault detection. To this end, the proposed methodology integrates Mahalanobis-distance-based sample weighting, adversarial feature adaptation, and dynamic confidence thresholding. Its effectiveness is evaluated through cross-condition transfer experiments using the CWRU bearing dataset and a laboratory motor-bearing platform, with particular emphasis on known-class classification, unknown-class detection, and overall diagnostic performance.

The remainder of this paper is organized as follows. Section 2 introduces the theoretical foundations of the proposed method, including Mahalanobis distance, dynamic threshold estimation, and weighted adversarial feature alignment. Section 3 describes the overall architecture of WAAN, its loss functions, and the diagnostic procedure. Section 4 presents the experimental settings and evaluates the proposed method using the CWRU bearing dataset and a laboratory motor-bearing platform, followed by an analysis of the diagnostic results. Finally, Section 5 summarizes the main findings and discusses the potential applications of the proposed approach.

2. Theoretical Analysis

2.1. Research Problem and Methodological Formulation

This paper investigates open-set unsupervised domain-adaptive fault diagnosis. We define a labeled source domain dataset $D^s = \left\{\left(x_i^s, y_i^s\right)\right\}_{i=1}^{n}$ where x_i^s represent the i-th sample and y_i^s represents the label of that sample, and n denote the number of source domain samples. We also define an unlabeled target domain dataset $D^t = \left\{x_j^t\right\}_{j=1}^{m}$ where x_j^t represent the j-th target domain data sample, and m denote the number of target domain samples. The label space of the source domain is a subset of the target domain's label space

$D^s \subseteq D^t$, meaning that the target domain contains faults that have never occurred in the source domain. Additionally, in unsupervised domain-adaptive diagnosis, the probability distributions of the source and target domains differ $P(x^s) \neq P(x^t)$, where $P(x^s)$ represents the probability distribution of the source domain data, and $P(x^t)$ represents the probability distribution of the target domain data.

Under the open-set domain adaptation setting, the objective is to learn transferable feature representations that distinguish shared fault categories while reducing the influence of unknown target-domain samples. The proposed WAAN framework addresses this problem through Mahalanobis-distance-based sample weighting and adaptive confidence thresholding. The former regulates the contribution of target samples to adversarial alignment, whereas the latter determines whether a target sample should be classified as a known category or rejected as unknown. The effectiveness of the framework is assessed through cross-condition transfer experiments on public and laboratory bearing datasets.

In open-set fault diagnosis, unknown fault detection and unknown fault identification represent two different levels of diagnostic capability. Unknown fault detection aims to determine whether a target-domain sample belongs to one of the known fault categories or should be rejected as unknown. In contrast, unknown fault identification, in a stricter sense, requires further distinguishing the specific fault types among previously unseen categories. The proposed WAAN framework focuses on unknown fault detection and rejection rather than fine-grained identification of unknown fault types. Specifically, all fault categories absent from the source domain are assigned to a unified unknown label during evaluation. Therefore, the reported unknown-class detection rate reflects the model's ability to distinguish unknown samples from known categories, rather than its ability to identify the specific physical types of unknown faults.

2.2. Mahalanobis Distance

Mahalanobis distance [14] is a distance metric based on the characteristics of data distribution. It calculates the "statistical distance" between a sample point and the data distribution by introducing a covariance matrix to adjust the correlation and scale differences between features. Mahalanobis distance is widely used in anomaly detection tasks [15,16]. If the distance exceeds a given threshold, it is judged as an anomaly.

2.3. Dynamic Threshold Estimation

To reduce the negative impact of unknown class samples in the target domain on the model classification, a dynamic threshold weighting strategy based on a sliding window is proposed [17,18]. For a sample in the target domain, the maximum predicted probability of its known class is calculated:

$$P_{max} = \max\left(\text{Softmax}\left(\hat{y}_{know}\right)\right) \tag{1}$$

where $\hat{y}_{know}$ is the logical output of the known category.

Maintain a sliding window queue of length W that stores historical confidence values $\left\{p_{max}^{(t)}\right\}$. The dynamic threshold τ is obtained by calculating the α-percentile of the confidence values in the queue:

$$\tau^{(t)} = \text{Percentile}\left(\left\{p_{max}^{(t-W+1)}, \ldots, p_{max}^{(t)}\right\}, 100\alpha\right) \tag{2}$$

The quantile parameter α determines the position of the rejection threshold within the recent prediction-confidence distribution, whereas the sliding-window length W specifies the number of confidence observations used for threshold estimation. A smaller W may enable faster adaptation to changes in confidence distributions but can result in greater

threshold fluctuations, while a larger W may provide a more stable estimate at the expense of slower adaptation. Importantly, α should not be interpreted as the actual proportion of unknown samples, since known and unknown samples may exhibit overlapping confidence distributions under domain shifts. Both α and W are therefore hyperparameters that must be specified without using target-test labels. In practical applications, a predefined configuration or a validation procedure independent of target-test labels is required, and the sensitivity of the method to these choices remains an important consideration.

2.4. Adversarial Feature Alignment

This paper proposes a dynamic weighted adversarial training mechanism based on uncertainty estimation [3,19] to improve the robustness of cross-domain feature alignment. The specific implementation consists of the following three steps:

The domain transferability of target domain samples is quantified by the deviation of the Mahalanobis distance distribution:

In the proposed Mahalanobis-distance-based weighting mechanism, the feature center of each known source-domain class is estimated using its corresponding labeled samples. Meanwhile, a shared global covariance matrix is constructed from all source-domain feature representations to characterize their overall statistical variation and feature correlations. Specifically, the global covariance matrix is estimated using the unbiased sample covariance estimator, as defined below:

$$\hat{\Sigma} = \frac{1}{N_s - 1} \sum_{i=1}^{N_s} (F(x_i^s) - \overline{\mu})(F(x_i^s) - \overline{\mu})^T \tag{3}$$

$$\overline{\mu} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathrm{F}(x_i^s) \tag{4}$$

Here, N_s denotes the number of source-domain samples, and $F\left(x_i^s\right)$ represents the extracted feature vector. The global covariance matrix is estimated from all source-domain features, whereas each known class maintains a separate feature center. To improve numerical stability and alleviate potential singularity or ill-conditioning, the estimated global covariance matrix was regularized by adding a diagonal term with $\epsilon = 10^{-5}$ before standard matrix inversion. The resulting inverse covariance matrix is shared across all known classes for Mahalanobis-distance calculation. For each target-domain sample, the Mahalanobis distances to the known-class centers are calculated separately, and the minimum distance is adopted to quantify its deviation from the known source-domain distributions.

$$\sum_{\mathrm{reg}} = \hat{\sum} + \epsilon \mathrm{I} \tag{5}$$

$$\mathrm{D}_{\mathrm{Mahal}}(\mathrm{x}^{\mathrm{t}}) = \min_{\mathrm{k}} \sqrt{\left(\mathrm{f}^{\mathrm{t}} - \mu_{\mathrm{k}}\right)^{\top} \Sigma^{-1} \left(\mathrm{f}^{\mathrm{t}} - \mu_{\mathrm{k}}\right)} \tag{6}$$

Larger distances indicate greater deviation from the known class distribution.

After normalizing the above metrics, the sample-level adversarial weights $\mathrm{w}(\mathrm{x}^{\mathrm{t}}) \in [0, 1]$ are generated.

$$\mathrm{w}(\mathrm{x}^{\mathrm{t}}) = 1 - \frac{\mathrm{D}_{\mathrm{Mahal}}(\mathrm{x}^{\mathrm{t}}) - \mathrm{D}_{\min}}{\mathrm{D}_{\max} - \mathrm{D}_{\min}} \tag{7}$$

The weight $\mathrm{w}(\mathrm{x}^{\mathrm{t}})$ reflects the likelihood that a sample belongs to a known class; the larger the value, the stronger the constraint on domain alignment during adversarial training.

When training a domain classifier, the adversarial loss for samples from the target domain is weighted:

$$\mathrm{L}_{\mathrm{adv}} = \mathbb{E}_{\mathrm{x} \in \mathrm{D}_{\mathrm{s}}}[\log \mathrm{D}(\mathrm{F}(\mathrm{x}))] + \mathbb{E}_{\mathrm{x} \in \mathrm{D}_{\mathrm{x}}}\left[\mathrm{w}\left(\mathrm{x}^{\mathrm{t}}\right) \cdot \log(1 - \mathrm{D}(\mathrm{F}(\mathrm{x})))\right] \tag{8}$$

3. Materials and Methods

For cross-domain open-set diagnosis of rolling bearings, the WAAN (Weighted Adversarial Adaptive Network) mainly consists of a feature extractor, an open-set classifier, and a weighted adversarial domain adaptation module. The network framework is shown in Figure 1. First, a Fast Fourier Transform is performed on the source and target domain samples to obtain frequency domain signals, which are then input into the feature extractor to extract high-dimensional features. These features are then input into the open-set classifier to obtain the classification loss, and into the weighted domain adversarial classifier to obtain the domain classification loss. The network parameters are updated using backpropagation based on the total loss. During the testing phase, the target domain samples are input into the trained feature extractor. If the highest predicted probability output by the classifier exceeds a dynamic threshold, it is determined to be a shared category; otherwise, it is considered an unknown fault. Through this process, the model achieves automatic intelligent fault diagnosis of bearings in cross-domain open-set scenarios.

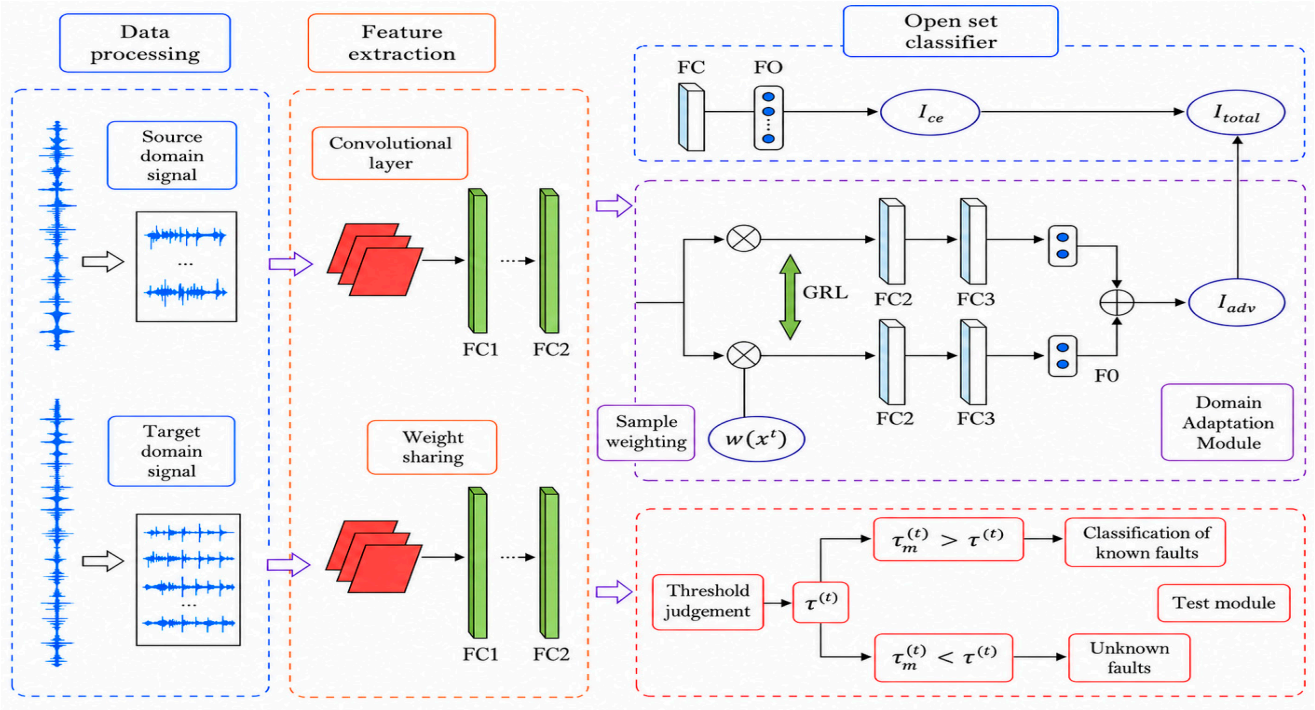


Figure 1. Fault diagnosis model based on weighted adversarial adaptive network.

3.1. Network Architecture

The WAAN consists of four core modules: the Feature Extractor, the Open-Set Classifier, the Domain Classifier, and the Metric Generator. The detailed structure of each module is summarized in Table 1.

The feature extractor employs a one-dimensional convolutional neural network [18,19] (1D-CNN), with the input being a frequency-domain vibration signal (length 1024) transformed by FFT. Each convolutional layer is followed by batch normalization, ReLU activation, and max-pooling layers. A Standout layer is incorporated after the fully connected layers to adaptively regulate feature activations and enhance feature representation robustness.

Open-set classifiers address the issue of unknown categories in the target domain. Building upon the known number of categories η_{source} in the source domain, open-set classifiers add an additional dimension $n_{source} + 1$ to represent the unknown category, meaning the

network's output dimension is $n_{sorce} + 1$. During the forward pass, all categories (including the unknown category) are modeled jointly through log-probability normalization:

$$\log P = \log_softmax(z),\ P \in \mathbb{R}^{n_{source}+1} \tag{9}$$

Table 1. Detailed information about the CNN structure.

Module	Layer Type	Input Dimensions	Output Size
Feature extractor	Conv1d	(n, 1, 1024)	(n, 16, 1024)
	BatchNorm1d	(n, 16, 1024)	(n, 16, 1024)
	MaxPool1d	(n, 16, 1024)	(n, 16, 512)
	Conv1d	(n, 16, 512)	(n, 32, 512)
	BatchNorm1d	(n, 32, 512)	(n, 32, 512)
	MaxPool1d	(n, 32, 512)	(n, 32, 256)
	Conv1d	(n, 32, 256)	(n, 64, 256)
	BatchNorm1d	(n, 64, 256)	(n, 64, 256)
	MaxPool1d	(n, 64, 256)	(n, 64, 128)
	Flatten	(n, 64, 128)	(n, 8192)
	Linear (fc1)	(n, 8192)	(n, 1024)
	Linear (fc2)	(n, 1024)	(n, 1024)
	Standout	(n, 1024)	(n, 1024)
Operating condition classifier	Linear (fc)	(n, 1024)	(n, k)
	Softmax	(n, k)	(n, k)
Domain classifier	GradientReverseModule	(n, 1024)	(n, 1024)
	Linear	(n, 1024)	(n, 512)
	Linear	(n, 512)	(n, 1)
Metric generator	Linear (fc1)	(n, 1024)	(n, 1000)
	BatchNorm1d	(n, 1000)	(n, 1000)
	Linear (fc3)	(n, 1000)	(n, k)

By setting a pre-defined confidence threshold, samples with a known class probability below the threshold are classified as unknown classes. This allows for the effective detection of novel faults (unknown classes) while maintaining accurate classification of known classes.

To address the distributional differences between the source and target domains, a domain adaptation strategy based on weighted adversarial learning was adopted [10,14]. The specific method involves constructing a domain classifier and connecting a gradient reversal layer [19] (GRL) to its front end to ensure that the feature extraction module receives gradient signals in opposite directions during the backpropagation phase, thereby enabling it to learn domain-invariant representational features. Mahalanobis distance is used to estimate the class centers of known categories in the source domain, measuring the distance between the target domain sample and each class center. Samples with greater distances are assigned smaller weights during adversarial learning. This constraint prevents unknown samples in the target domain from being forcibly matched with any source domain sample, avoiding the misclassification of unknown faults into known categories.

During the testing phase, the open-set classifier models all categories (including the unknown class) as part of the same probability space using log_softmax, and determines the unknown class based on the probability values:

$$\hat{y}(x_t) = \begin{cases} \text{argmax} P_k(x_t), & P_{max}(x_t) \geq \tau^{(t)}, \\ \text{unknown}, & P_{max}(x_t) < \tau^{(t)} \end{cases} \tag{10}$$

where $P_{max}(x_t)$ denotes the maximum predicted probability among all known source-domain classes, and $\tau^{(t)}$ represents the dynamically updated confidence threshold. If $P_{max}(x_t)$ is greater than or equal to $\tau^{(t)}$, the target sample is classified as the known category corresponding to the highest predicted probability. Otherwise, the sample is rejected as belonging to an unknown fault category. This decision rule enables the model to distinguish between known and potentially unknown fault samples according to their prediction confidence.

3.2. Loss Function Formulation and Optimization

The total network loss consists of two components:

(1). Source-domain supervision loss: For source-domain samples, the cross-entropy loss is calculated using only the known class probabilities, where k is the number of source domain categories and y_k^s is the true label of the source domain:

$$L_{CE} = -\sum_{k=1}^{K} y_k^s \log p_k \tag{11}$$

(2). Domain adversarial loss: This loss is applied to the output of the domain classifier and the domain labels (1 for the source domain and 0 for the target domain). When combined with GRL, this loss drives the feature extractor to generate domain-agnostic features. The domain classification loss is used to optimize the consistency between the feature distributions of the source and target domains. The total loss function of the network is:

$$\mathcal{L}_{Total} = \mathcal{L}_{CE} - \lambda L_{adv} \tag{12}$$

(3). Where λ is the loss balance coefficient, which was determined to be 0.2 through experiments.

Based on the total loss function of the network, the weighted adversarial adaptive network is optimized and trained by calculating the backpropagation. Taking Equation (12) as the optimization objective, an open set classifier with accurate classification ability is obtained by minimizing the source domain supervised fault classification error, while maximizing the domain discriminator error to extract domain shared features, and finally forming the optimization method shown in Equation (13).

$$L(\theta_f^*, \theta_c^*, \theta_D^*) = \min_{\theta_f, \theta_c, \theta_D} [\mathcal{L}_{CE}(\theta_f, \theta_c) - \lambda \mathcal{L}_{adv}(\theta_f, \theta_D)] \tag{13}$$

Based on Equation (13) and SGD algorithm, the parameters θ_f, θ_c, and θ_D are updated using the formula, where ϵ is the learning rate.

$$\theta_f \leftarrow \theta_f - \epsilon \left(\frac{\partial L_{ce}}{\partial \theta_f} - \lambda \frac{\partial \mathcal{L}_{adv}}{\partial \theta_f} \right) \tag{14}$$

$$\theta_c \leftarrow \theta_c - \epsilon \frac{\partial \mathcal{L}_{ce}}{\partial \theta_c} \tag{15}$$

$$\theta_D \leftarrow \theta_D - \epsilon \frac{\partial \mathcal{L}_{adv}}{\partial \theta_D} \tag{16}$$

3.3. Diagnostic Procedure of the Proposed Model

Figure 2 shows the fault diagnosis process for WAAN, which includes the following steps:

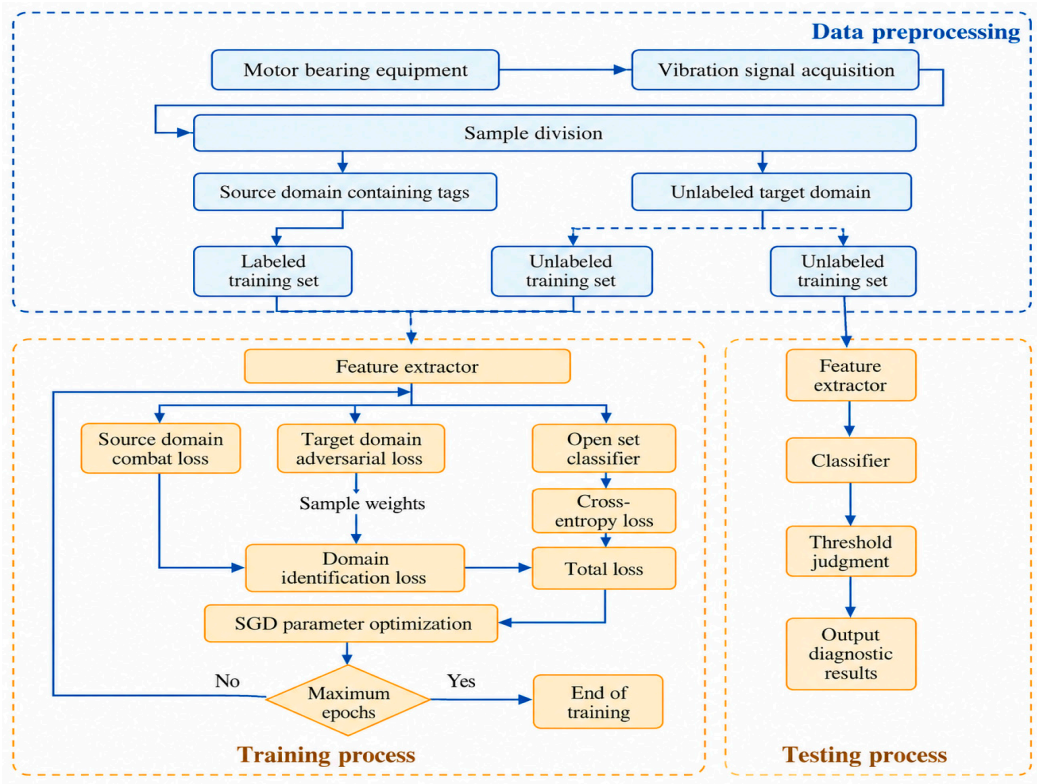


Figure 2. WAAN diagnostic flowchart.

(1). Before training begins, use a feature extractor to extract features from the source domain data and calculate the feature mean (class center) for each class. Calculate the covariance matrix and its inverse matrix for all source domain features, which are used for subsequent Mahalanobis distance calculation.

(2). The source and target domain data are input into the feature extractor to obtain the corresponding feature representations $f_s = F(x_s)$, $f_t = F(x_t)$. The extracted features are then used by classifier C to obtain the predicted classification outputs $y_s = C(f_s)$, $y_t = C(f_t)$ for all categories. Meanwhile, the domain classifier D is used to estimate prediction probabilities for both source and target samples.

(3). Calculate the classification loss $\mathcal{L}_{CE}$ of the model on the source domain sample (x_s,y_s) according to Equation (11). Apply *log_softmax* to the prediction result of the target domain according to Equation (9) to obtain the prediction probability P_{max} of the largest known class.

(4). The Mahalanobis distance of the target domain data sample distribution offset is calculated according to Equation (6). The greater the distance, the greater the uncertainty. The distance value is normalized and the weight is calculated according to Equation (7) to measure the importance of the sample. According to Equation (8), for source domain data, the binary cross-entropy loss between its domain prediction and the true label is calculated; for target domain data, the binary cross-entropy loss between its domain prediction and the true label is calculated and multiplied by the corresponding weight. The classification loss and adversarial loss are weighted and summed to obtain the total loss, Equation (12).

(5). Backpropagation is performed based on the total loss in Equation (12), and the model parameters are updated based on Equations (14)–(16).

During inference, the diagnostic procedure consists of signal acquisition, preprocessing, feature extraction, classification, and dynamic threshold-based rejection. Therefore, the overall detection latency includes signal acquisition time, computational inference time, and threshold updating overhead. These components should be distinguished when evalu-

ating the real-time applicability of the proposed framework. Since direct inference-time and end-to-end latency measurements were not included in the present experiments, real-time diagnostic performance has not yet been quantitatively established.

4. Experimental Results and Discussion

During model training, stochastic gradient descent (SGD) was selected as the optimization strategy. Specific parameters were set as follows: the initial learning rate was set to 0.0002 to ensure stable model updates; the momentum factor was set to 0.9 to improve convergence speed; and the weight decay coefficient was set to 0.0005 to suppress the risk of overfitting.

The sliding-window length was set to W = 400 as a predefined empirical configuration to balance threshold stability and responsiveness to changes in prediction confidence. Smaller windows may produce greater threshold fluctuations, whereas larger windows may delay adaptation to distribution shifts. The influence of the quantile parameter α was examined in Figure 3, where the best-performing values were 0.5 and 0.4 for the balanced and class-imbalanced tasks, respectively. These results indicate task-dependent sensitivity rather than universally optimal parameter settings. A preliminary sensitivity analysis of the sliding-window length W is presented in Section 4.2. However, the joint sensitivity of W and α, together with a validated label-free parameter selection strategy, remains to be investigated.

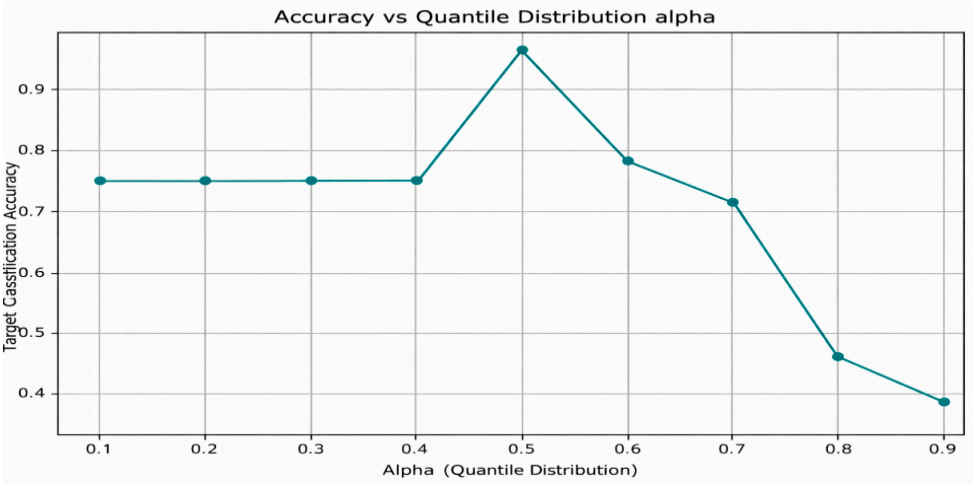


Figure 3. The diagnostic accuracy of WAAN under different threshold quantiles for four classification tasks was investigated.

Sudden changes in operating conditions may cause temporary shifts in prediction-confidence distributions, potentially affecting the stability of the dynamic threshold. In addition, an increase in the proportion of unknown samples may alter the empirical confidence distribution and consequently influence threshold estimation. Although the sliding-window mechanism updates the threshold using recent confidence observations, its transient response under abrupt load changes and rapidly varying unknown-class proportions has not been quantitatively evaluated in this study.

4.1. Experimental Data Setup

Case Study 1: Case Western Reserve University (CWRU) Bearing Dataset.

A publicly available dataset provided by the bearing testing platform at Case Western Reserve University was selected. This dataset covers four typical operating states: healthy condition, inner ring damage, rolling element defects, and outer ring damage. The test bearings were divided into three different operating conditions (0 hp load, 1 hp load, and 2 hp load) based on different loads, corresponding to speeds of 1797, 1772, and 1750 r/min,

respectively. Six migration tasks were set up as shown in Table 2. Taking the first migration task as an example, 1797–1772 indicates that the operating condition at speed 1797 is selected as the source domain, and the operating condition at speed 1772 is selected as the target domain. The source domain consists of normal conditions and two types of data: outer ring defects and rolling element defects; the target domain consists of normal conditions and three other defect types.

Table 2. CWRU Different load migration experiments.

Serial Number	Migration Task	Source	Target Domain
1	1797–1772	N/OF/RF	N/IF/OF/RF
2	1797–1750		
3	1772–1750		
4	1772–1797		
5	1750–1797		
6	1750–1772		

To prevent potential data leakage from highly correlated vibration samples, the original recordings were partitioned before sliding-window segmentation. All segments derived from the same independent recording were assigned exclusively to the corresponding data partition, preventing overlapping segments from appearing in both training and test sets. The same data partitioning protocol was applied to all compared models to ensure evaluation consistency.

4.2. Experimental Results and Analysis

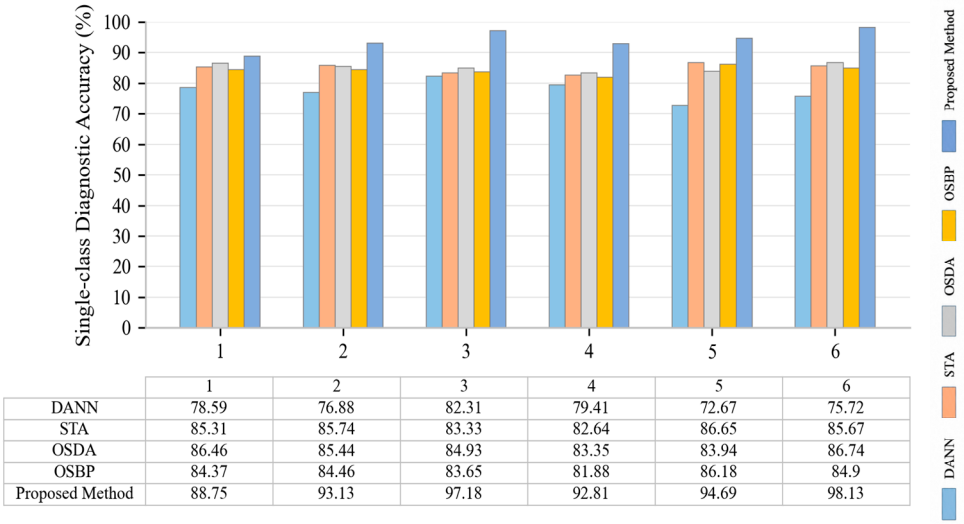
Figure 3 illustrates the influence of the quantile parameter α on the diagnostic accuracy of WAAN for the investigated transfer task. The highest accuracy was observed at $\alpha = 0.5$, although unknown samples accounted for 25% of the target-domain data. This observation indicates that the most suitable confidence quantile does not necessarily correspond to the proportion of unknown samples. The dynamic threshold is determined by the distribution of prediction confidence rather than by the actual unknown-class proportion. In particular, unknown samples may receive relatively high confidence scores, while certain known samples may exhibit reduced confidence under distribution shifts. Such overlap can influence the trade-off between unknown-class rejection and known-class preservation, resulting in different performance across quantile settings. Therefore, the value $\alpha = 0.5$ should be regarded as the best-performing setting among those examined for this specific task rather than a theoretically optimal or universally applicable parameter.

To investigate the influence of the sliding-window length on the dynamic threshold mechanism, five candidate values of W (100, 200, 400, 600, and 800) were evaluated while keeping the trained model, quantile parameter α, and other inference settings unchanged. As shown in Table 3, the reported values indicate that W = 400 achieves the highest known-class accuracy, unknown-class recall, and H-score among the examined configurations. Smaller windows may produce greater fluctuations in confidence-based threshold estimation, whereas larger windows may reduce responsiveness to changes in the target-domain confidence distribution. These findings suggest that an intermediate window length can provide a favorable balance between threshold stability and adaptability under the investigated conditions. Nevertheless, this result does not establish W = 400 as a universally optimal configuration.

Table 3. Sensitivity Analysis of the Sliding-Window Length W.

W	Known Accuracy (%)	Unknown Recall (%)	H-Score (%)
100	98.75	69.50	81.58
200	99.17	73.25	84.26
400	99.72	76.25	86.42
600	99.58	74.50	85.23
800	99.50	71.50	83.21

To evaluate the effectiveness of the proposed WAAN framework, comparative experiments were conducted against four representative domain adaptation methods, namely DANN, OSDA, STA, and OSBP, under six cross-condition transfer tasks. The diagnostic accuracy of each method is presented in Figure 4.



	1	2	3	4	5	6
DANN	78.59	76.88	82.31	79.41	72.67	75.72
STA	85.31	85.74	83.33	82.64	86.65	85.67
OSDA	86.46	85.44	84.93	83.35	83.94	86.74
OSBP	84.37	84.46	83.65	81.88	86.18	84.9
Proposed Method	88.75	93.13	97.18	92.81	94.69	98.13

Figure 4. Comparison of diagnostic accuracy under different methods (From left to right, the colors are light blue, light orange, light gray, yellow, and blue-purple, corresponding to DANN, STA, OSDA, OSBP, and Proposed Method, respectively, from top to bottom in the table.).

As shown in Figure 4, WAAN achieves the highest diagnostic accuracy across the six investigated transfer tasks, outperforming DANN by an average margin of approximately 16 percentage points. This improvement can be attributed to the complementary roles of Mahalanobis-distance-based sample weighting and dynamic thresholding. The former reduces the influence of potentially unknown target samples during adversarial feature alignment, whereas the latter adaptively adjusts the rejection criterion according to recent prediction-confidence statistics. Together, these mechanisms contribute to more reliable cross-domain fault diagnosis under the investigated open-set conditions.

To provide a comprehensive evaluation of open-set diagnostic performance, additional metrics, including unknown-class precision, recall, F1-score, and H-score, were introduced. The unknown category was treated as the positive class, while all known categories were treated as negative classes when calculating unknown-class detection metrics. The unknown-class precision, recall, F1-score, and H-score are defined as follows:

$$P_u = \frac{TP_u}{TP_u + FP_u} \tag{17}$$

$$R_u = \frac{TP_u}{TP_u + FN_u} \tag{18}$$

$$Fl_u = \frac{2P_u R_u}{P_u + R_u} \tag{19}$$

$$H = \frac{2A_k R_u}{A_k + R_u} \tag{20}$$

where TP_u denotes the number of unknown samples correctly detected as unknown, FP_u represents known samples incorrectly rejected as unknown, and FN_u denotes unknown samples incorrectly classified as known. A_k represents the classification accuracy of known-class samples, calculated as the proportion of known samples assigned to their correct known categories. The H-score is the harmonic mean of known-class accuracy and unknown-class recall, providing a balanced evaluation of open-set diagnostic performance.

Figure 5 presents the confusion matrices of the six open-set transfer tasks. Labels 0–2 represent the shared fault categories, including the normal condition, outer race faults, and rolling element faults, whereas label 3 represents unknown inner race faults. To comprehensively assess known-class classification and unknown-class detection, additional evaluation metrics are reported in Table 4.

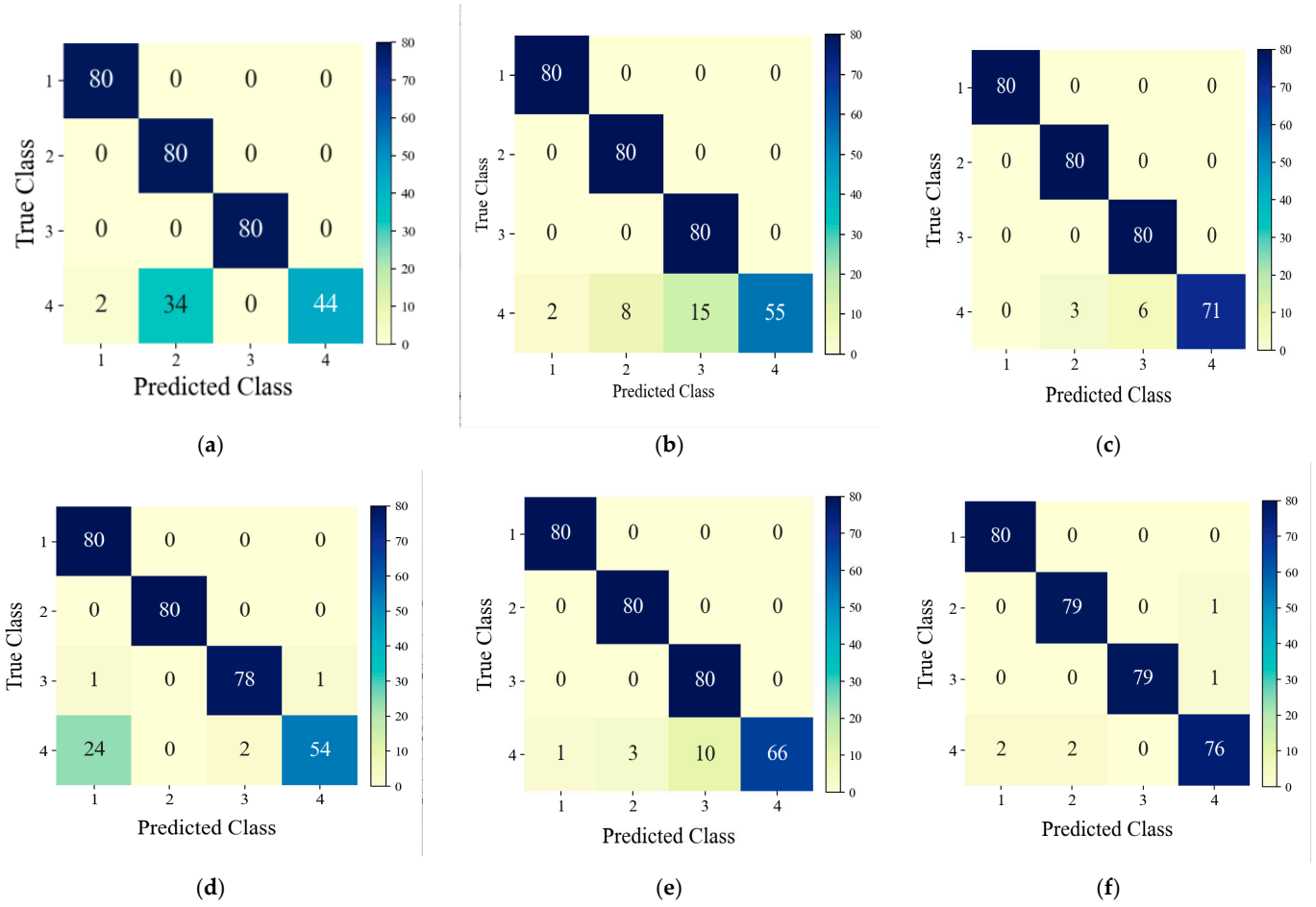


Figure 5. The proposed method is used to visualize the confusion matrix of diagnostic results. (**a**) Description of Experiment 1; (**b**) Description of Experiment 2; (**c**) Description of Experiment 3; (**d**) Description of Experiment 4; (**e**) Description of Experiment 5; (**f**) Description of Experiment 6.

Table 4. Open-set diagnostic performance of WAAN across six transfer tasks.

Task	Known Acc (%)	Unknown Precision (%)	Unknown Recall (%)	Unknown F1 (%)	H-Score (%)
1	100.00	100.00	55.00	70.97	70.97
2	100.00	100.00	68.75	81.48	81.48
3	100.00	100.00	88.75	94.04	94.04
4	99.17	98.18	67.50	80.00	80.33
5	100.00	100.00	82.50	90.41	90.41
6	99.17	97.44	95.00	96.20	97.04
Pooled	99.72	99.19	76.25	86.22	86.42

Note: The pooled metrics were calculated by aggregating the confusion matrices of all six transfer tasks before computing the evaluation metrics.

The detailed open-set evaluation results for the six transfer tasks are summarized in Table 4.

As summarized in Table 4, among the 1440 known-class samples across six transfer tasks, 1436 were correctly classified, yielding a pooled known-class accuracy of 99.72%. Of the 480 unknown-class samples, 366 were correctly detected, corresponding to an unknown-class recall of 76.25%. Only three known-class samples were incorrectly rejected as unknown, resulting in an unknown-class precision of 99.19%. The pooled unknown-class F1-score and H-score were 86.22% and 86.42%, respectively.

These results reveal an important difference between known-class classification and unknown-class detection. Although WAAN maintains high known-class accuracy and rarely rejects known samples incorrectly, 114 unknown samples were assigned to known categories across the six tasks. In particular, Task 1 achieved an unknown-class recall of only 55.00%, indicating that unknown-fault detection remains challenging in certain transfer scenarios. Thus, the high overall classification accuracy should not be interpreted as uniformly strong performance across known and unknown categories. The additional precision, recall, F1-score, and H-score metrics provide a more comprehensive assessment of this trade-off.

It should be noted that the existing comparison with DANN, STA, OSDA, and OSBP mainly evaluates the overall diagnostic performance of different domain adaptation approaches. It does not isolate the effectiveness of the proposed unknown-class rejection mechanism relative to alternative out-of-distribution detection strategies. In particular, maximum-softmax-probability detection and Mahalanobis-distance-based OOD detection use different scoring criteria and may exhibit different trade-offs between known-class preservation and unknown-class rejection. Therefore, the present results do not establish that WAAN outperforms these alternative OOD detection methods. A controlled comparison using identical feature representations, target samples, and label-independent threshold calibration would be necessary to establish such an advantage. WAAN is trained by assigning different weights to different samples to suppress the interference of unknown categories in domain alignment. Figure 6 shows the changes in the weights of the target domain samples during training. It can be seen that there is a significant difference in weights between known and unknown faults, indicating that the proposed method can effectively distinguish and suppress the alignment of unknown faults.

Although Mahalanobis-distance-based weighting helps characterize the distributional deviation of target samples, its reliability depends on the representativeness of the source-domain feature distributions. When known and unknown fault categories exhibit overlapping feature distributions, potentially unknown samples may receive relatively high alignment weights. Conversely, substantial changes in operating conditions may cause known target samples to deviate from their corresponding source-domain class centers and receive reduced weights. These effects may reduce the effectiveness of selective domain

alignment and contribute to unknown-class misclassification. Therefore, the robustness of the weighting mechanism under severe feature overlap and larger domain shifts requires further validation.

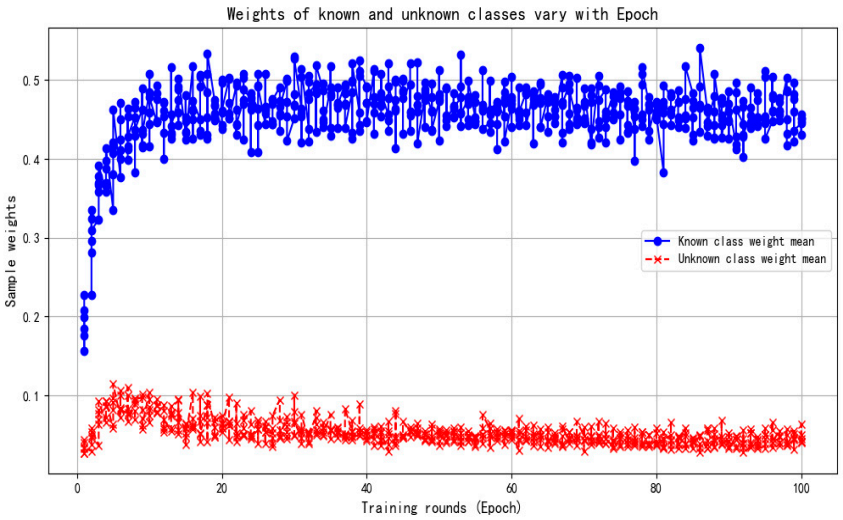


Figure 6. Weight variations in known and unknown category samples during training.

Ablation Study of the Proposed WAAN.

To further investigate the individual and combined contributions of the proposed mechanisms, an ablation study was designed based on the six CWRU cross-condition transfer tasks. Five model variants were considered: CNN + Fixed, DANN + Fixed, WAAN + Fixed, DANN + Dynamic, and Full WAAN. These variants were constructed by selectively enabling or disabling adversarial domain adaptation, Mahalanobis-distance-based sample weighting, and dynamic thresholding. Specifically, CNN + Fixed served as the baseline without adversarial adaptation or sample weighting; DANN + Fixed introduced adversarial domain adaptation; WAAN + Fixed further incorporated Mahalanobis-distance-based weighting; DANN + Dynamic combined adversarial adaptation with dynamic thresholding; and Full WAAN integrated all three components. The same feature extraction architecture and training configurations were adopted to ensure comparability among the model variants. Known-class accuracy, unknown-class precision, recall, F1-score, and H-score were selected as the evaluation metrics.

Table 5 provides a component-wise comparison of the five WAAN variants. Based on the tabulated values, Full WAAN achieves the highest known-class accuracy (99.72%), unknown-class recall (76.25%), and H-score (86.42%). Its H-score is 7.61 percentage points higher than that of WAAN + Fixed and 5.81 percentage points higher than that of DANN + Dynamic. These numerical comparisons illustrate the expected complementary contributions of sample weighting and dynamic thresholding to balancing known-class classification and unknown-fault detection.

Class Imbalance Experiment.

Considering the varying probabilities of different equipment failure types in actual production environments, leading to an imbalance in the number of samples collected across different data categories, and the potential for multiple types of unknown failures in the open set, fault diagnosis faces significant challenges. To demonstrate the diagnostic performance of the proposed method on imbalanced unknown samples, relevant experiments were designed. An imbalanced dataset was created using nine types of failures under four load conditions from the CWRU dataset, along with a set of health data. Six states were selected as known categories, leaving four categories—0.1778 mm inner race failure, 0.3556 mm rolling element failure, 0.3556 mm outer race failure, and 0.5334 mm

rolling element failure—as unknown categories. In this case, the number of samples in the unknown categories is four times the number of samples in the known categories. The data classification is shown in Table 6.

Table 5. Ablation study of the principal components of WAAN across six CWRU transfer tasks.

Model	Known Acc (%)	Unknown Precision (%)	Unknown Recall (%)	Unknown F1 (%)	H-Score (%)	Acc (%)
CNN + Fixed	96.20	75.00	43.50	55.05	59.91	83.03
DANN + Fixed	98.10	82.00	52.80	64.23	68.64	86.78
WAAN + Fixed	99.10	91.30	65.40	76.21	78.81	90.68
DANN + Dynamic	98.50	91.90	68.20	78.30	80.61	90.93
Full WAAN	99.72	99.19	76.25	86.22	86.42	93.85

Table 6. Unbalanced dataset partitioning.

Fault Type	Degree of Damage	Fault Label	Source Domain (Training Set)	Target Domain (Training Set)	Target Domain (Test Set)	Category
N	-	0	120	120	80	Known class
IF	0.1778 mm	6	0	120	80	Unknown class
BF	0.1778 mm	1	120	120	80	Known class
OF	0.1778 mm	2	120	120	80	Known class
IF	0.3556 mm	3	120	120	80	Known class
BF	0.3556 mm	6	0	120	80	Unknown class
OF	0.3556 mm	6	0	120	80	Unknown class
IF	0.5334 mm	4	120	120	80	Known class
BF	0.5334 mm	6	0	120	80	Unknown class
OF	0.5334 mm	5	120	120	80	Known class

Figure 7 presents the effect of the quantile parameter α on the diagnostic accuracy of WAAN under the class-imbalanced 1730–1750 r/min transfer task. The diagnostic accuracy increased as α varied from 0.1 to 0.4 and subsequently decreased when α exceeded 0.4, with the highest accuracy observed at $\alpha = 0.4$. Compared with the result shown in Figure 3, this finding suggests that the most suitable quantile may vary with the characteristics of the target-domain confidence distribution, including class imbalance and the overlap between known and unknown samples. A larger α generally increases the rejection threshold, which may improve unknown-class detection but also increase the risk of rejecting known samples. Conversely, a smaller α may preserve known-class predictions while allowing more unknown samples to be classified as known. Therefore, the observed results demonstrate task-dependent sensitivity to α rather than establishing a universally optimal quantile or proving improved robustness with increasing class numbers.

It should be emphasized that the quantile sweeps in Figures 3 and 7 are intended to analyze the influence of α on diagnostic performance, and the corresponding best-performing values were identified retrospectively from the experimental results. Such task-specific optimal values should not be interpreted as parameters that can be directly selected in a fully unlabeled target-domain deployment. In a practical unsupervised setting, α and W need to be fixed in advance using a validation protocol that excludes target-test labels. The present study does not establish a universally optimal label-free selection strategy, and a comprehensive joint sensitivity analysis of α and W has not been conducted. Therefore, parameter selection remains a limitation of the proposed dynamic threshold mechanism.

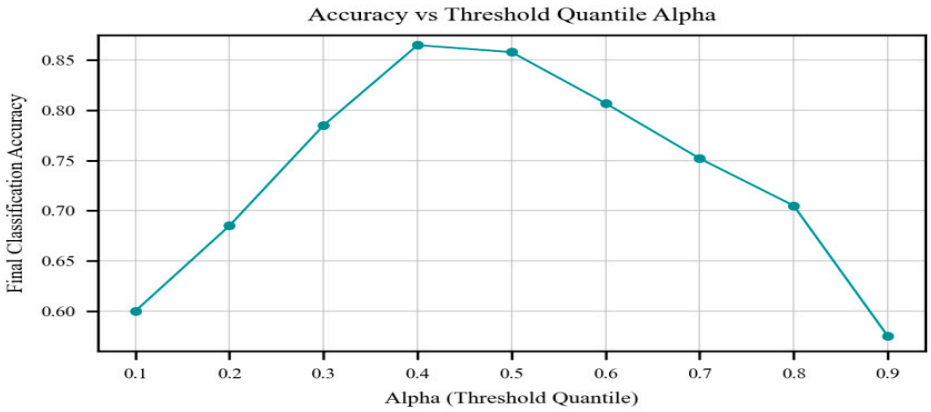
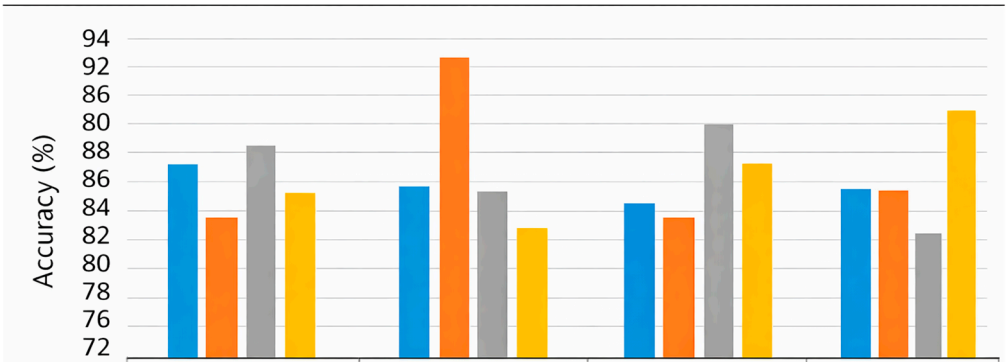


Figure 7. The diagnostic accuracy of WAAN under different threshold quantiles for seven classification tasks was investigated.

Although the sensitivity analysis in Table 3 provides an initial assessment of different sliding-window lengths, the joint effects of W and α under varying proportions of unknown samples have not been systematically evaluated. Furthermore, the transient behavior of the dynamic threshold under sudden operating-condition changes remains to be quantitatively investigated.

In the experiment with imbalanced unknown samples, a total of 16 open-set fault diagnosis routes were designed, as detailed in Figure 8, including 12 transfer diagnosis paths and 4 non-transfer paths. When the operating states of the target domain and the source domain are consistent, the model is trained using only samples under the same operating conditions. For example, path 1730–1730 represents a diagnosis experiment under the same operating conditions; conversely, if the operating conditions of the source domain and the target domain are different, the model is trained by combining labeled source domain data and unlabeled target domain data.



	T-1730	T-1750	T-1772	T-1797
S-1730	85.25	83.63	81.88	83.38
S-1750	81.38	92.5	80.75	80.32
S-1772	86.62	82.88	87.87	85.13
S-1797	82.88	80.12	85.13	88.63

Figure 8. Unbalanced experimental results of unknown category samples.

As shown in Figure 8, the average diagnostic accuracy under the same operating condition reached 88.56%; in the 1750 RPM condition, the diagnostic accuracy for non-transfer paths reached 92.5%. Training sets were established based on different operating conditions (1730, 1750, 1772, 1797), and other operating conditions were tested. The diagnostic accuracy for all transfer paths remained above 80%. Overall, the model exhibits

high diagnostic accuracy under class imbalance conditions, indicating its good open-set recognition capability.

To further investigate the performance of WAAN under class-imbalanced open-set conditions, Figure 9 presents the confusion matrix for the 1730–1750 r/min transfer task. The known-class classification accuracy was 87.91%, whereas the unknown-class detection rate was 75.63%, with 242 out of 320 unknown samples correctly detected. These results indicate that the performance gap between known-class classification and unknown-class detection persists under imbalanced conditions. In addition, the confusion matrix reveals misclassifications between known and unknown categories, suggesting that overlapping feature representations can lead to both missed unknown faults and incorrect rejection of known samples. Compared with the balanced open-set tasks, the class-imbalanced results further highlight the importance of evaluating the two types of diagnostic errors separately. Therefore, known-class accuracy and unknown-class recall should be considered jointly when assessing the reliability of WAAN in practical open-set applications.

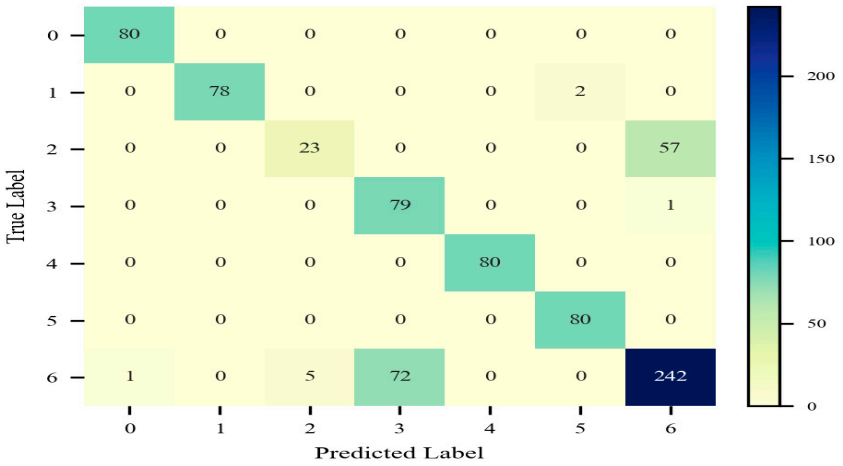


Figure 9. 1730–1750 Visualization of experimental results confusion matrix.

To visualize the feature clustering capability of the WAAN model, t-SNE technology was used. Taking the 1730–1750 experiment as an example, the features were visualized. The t-SNE visualization is shown in Figure 10. Purple represents the four unknown categories in the target domain. Among them, the purple samples close to the green samples are misidentified unknown category samples. The orange samples in the middle of the purple samples are 0.1778 mm outer ring faults that were misidentified as unknown samples. Apart from the above-mentioned clusters being close to the misclassified samples, the remaining samples maintain a relatively large clustering distance from samples of other different categories, indicating that the model can reliably cluster samples of the same state in the source and target domains that are known to be in the same category, even under the interference of unknown categories.

During training, the core idea of WAAN is to assign lower weights to unknown class samples to suppress the alignment of unknown samples with known state samples, thereby reducing the similarity of unknown samples. As shown in Figure 11, the average weights of samples are classified by fault type. During training, the average weights of known faults and unknown faults differ significantly. The average weight of known faults is between 0.2 and 1.0, while the average weight of unknown faults is less than 0.2, and it gradually decreases during the first 200 iterations.

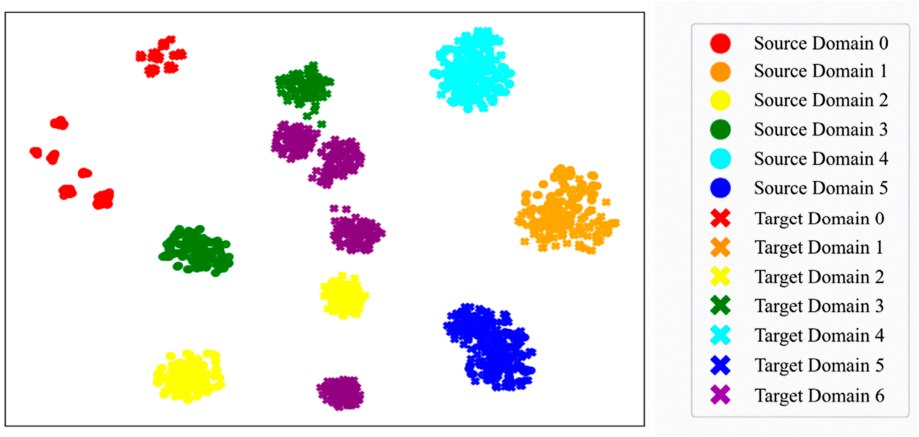


Figure 10. 1730–1750 Experimental t-SNE feature visualization results.

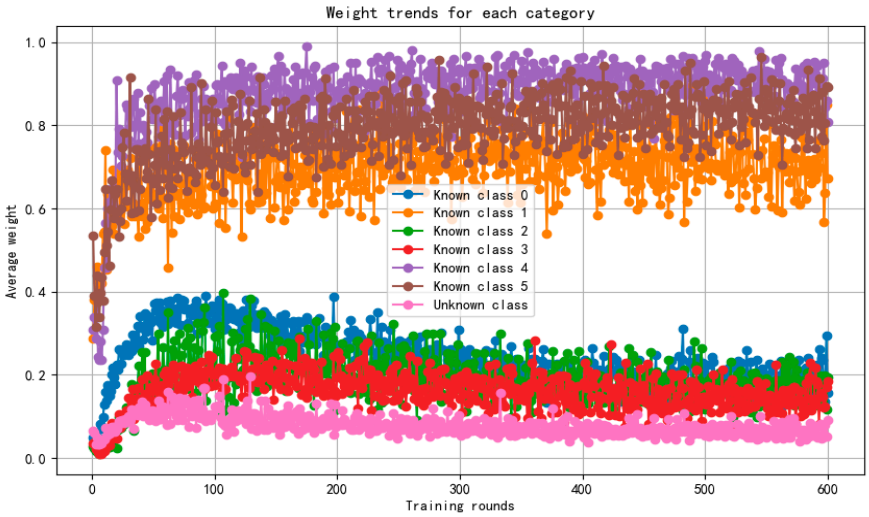


Figure 11. The average weight of each type of sample in the training.

Case Study 2: Actual Verification of the Motor Experimental Platform.

To further evaluate the effectiveness of the proposed WAAN method under variable operating conditions, experiments were conducted on a laboratory motor-bearing fault diagnosis platform. The experimental system consisted of a variable-frequency three-phase induction motor, an eddy-current brake, a bearing test assembly, an excitation control unit, and a vibration data acquisition system. The driving motor was a YVF2-90L-4 three-phase induction motor with a rated power of 1.5 kW, a rated voltage of 380 V, a rated current of 3.58 A, and a rated torque of 10 N·m. The motor supported variable-frequency operation, with a constant-torque frequency range of 5–50 Hz and a constant-power frequency range of 50–100 Hz. A WZ-20 eddy-current brake was employed as the loading device, with a rated power of 3 kW, a rated torque of 20 N·m, and a specified input speed range of 300–3000 r/min. The excitation control system was used to regulate the braking load, enabling the investigation of bearing fault diagnosis under different operating conditions.

Four bearing health states were investigated: normal condition (N), inner race fault (IF), outer race fault (OF), and rolling element fault (RF). Faulty bearing components with localized defects were employed to simulate different bearing damage conditions. The outer race fault specimen contained a localized defect on the raceway, with a nominal defect size of 0.8 mm, as indicated by the specimen marking. The rolling element fault specimen was also marked with a nominal defect size of 0.8 mm. These fault conditions, together with the normal bearing condition and the inner race fault condition, were used to construct the experimental fault diagnosis dataset.

During the experiments, the motor-bearing system was operated under four loading conditions of 0, 0.5, 1.0, and 1.5 kW. The experimental platform was equipped with a motor monitoring and loading control system, a data acquisition device, and computer-based monitoring software. Vibration signals were collected from the drive end and segmented using an overlapping sliding-window strategy with a window length of 1024 data points and a step size of 512 data points. For each health condition under each loading condition, 200 signal segments were selected, resulting in 800 samples per operating condition.

To construct the open-set diagnosis scenarios, dataset A corresponded to the no-load condition, while datasets B, C, and D corresponded to loading conditions of 0.5, 1.0, and 1.5 kW, respectively. Two source-domain subsets were established from dataset A. Dataset A1 contained normal, outer race fault, and rolling element fault samples, with inner race faults treated as the unknown category. Dataset A2 contained normal, inner race fault, and rolling element fault samples, with outer race faults treated as the unknown category.

Accordingly, six cross-condition transfer tasks were designed: A1 → B, A2 → B, A1 → C, A2 → C, A1 → D, and A2 → D. For each task, the labeled source-domain samples and 60% of the unlabeled target-domain samples were used for training, while the remaining 40% of target-domain samples were reserved for testing. The detailed dataset configurations and transfer tasks are presented in Tables 7 and 8, respectively. The experimental motor-bearing test platform and the corresponding fault specimens are shown in Figure 12.

Table 7. Fault data of different loads on the experimental platform.

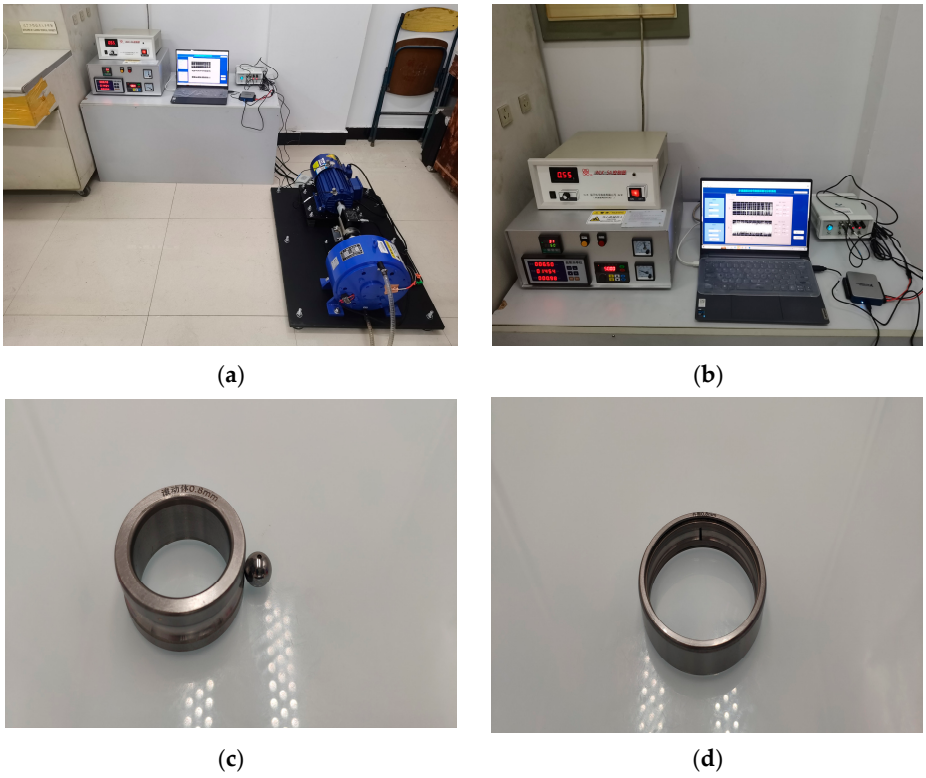
Dataset	Health Status	Sample Size	Load (kW)
A	N/IF/OF/RF	800	0
A1	N/OF/RF	600	0
A2	N/IF/RF	600	0
B	N/IF/OF/RF	800	0.5
C	N/IF/OF/RF	800	1
D	N/IF/OF/RF	800	1.5

Table 8. Different load migration experiments.

Experimental Task	Source	Target Domain	Training Data	Test Data
A1 → B	Dataset A1	Dataset B	Labeled source domain data and 60% unlabeled target domain data	The remaining 40% of the target domain data is unlabeled.
A2 → B	Dataset A2	Dataset B		
A1 → C	Dataset A1	Dataset C		
A2 → C	Dataset A2	Dataset C		
A1 → D	Dataset A1	Dataset D		
A2 → D	Dataset A2	Dataset D		

Based on the experimental tasks listed in Table 8, diagnostic accuracy was selected as the primary evaluation metric. Each transfer task was independently repeated ten times using different random seeds. The diagnostic performance was summarized using the mean and standard deviation across the ten runs to evaluate both average accuracy and experimental stability. Consistent data partitions and evaluation settings were maintained across the compared methods to ensure a fair comparison. The corresponding diagnostic results are presented in Figure 13. WAAN was compared with four representative baseline methods, namely DANN, OSDA, STA, and OSBP. The comparison shows that the highest diagnostic accuracy of the DANN model was 78.74%, with an average of 76.45%; the highest diagnostic accuracy of OSDA was 87.63%, with an average of 85.23%; the highest accuracy of STA was 88.61%, with an average of 85.67%; the highest diagnostic accuracy of OSBP was

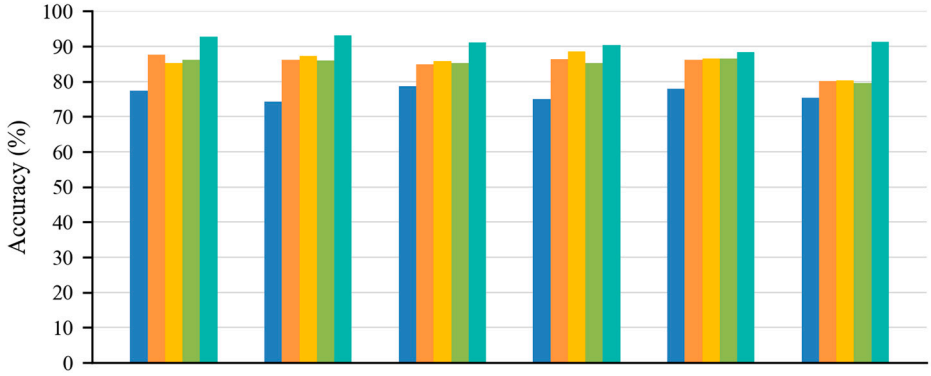
86.51%, with an average diagnostic accuracy of 84.79%; and the proposed method had an average diagnostic accuracy of 91.17% and a highest diagnostic accuracy of 93.06%. These results indicate that WAAN achieves competitive diagnostic performance on the investigated laboratory motor-bearing platform, supporting its applicability to the evaluated cross-load open-set diagnosis tasks. Nevertheless, its effectiveness under substantially larger domain shifts and across other types of machinery remains to be established.



(**a**) (**b**)

(**c**) (**d**)

Figure 12. Experimental motor-bearing test platform and fault specimens: (**a**) motor and loading system; (**b**) control and data acquisition system; (**c**) rolling element fault specimen; and (**d**) outer race fault specimen.



	A1→B	A2→B	A1→C	A2→C	A1→D	A2→D
DANN	77.43	74.33	78.74	74.98	77.89	75.32
OSDA	87.63	86.1	84.96	86.33	86.23	80.19
STA	85.34	87.36	85.76	88.61	86.58	80.34
OSBP	86.17	85.97	85.26	85.18	86.51	79.63
Proposed Method	92.84	93.06	91.13	90.31	88.43	91.25

Figure 13. Diagnostic accuracy comparison of different methods on the laboratory motor-bearing platform (The colors in the bar chart, from left to right, correspond to the methods in the table, from top to bottom.).

5. Conclusions

Current unsupervised domain adaptive diagnostic methods typically assume that the source and target domains contain the same health state categories. However, in practical applications, the test set often contains not only health state types seen during training but also fault mode samples not present in the training set. To address this issue, this paper proposes a weighted adversarial adaptive network (WAAN). By designing a target domain sample weighting strategy, during the adaptation process, only the domain alignment of health state samples shared between the source and target domains is enhanced, while the domain alignment of additional health state samples in the target domain with the source domain is weakened by using lower weights. In the testing phase of target domain fault classification, a pre-trained feature extractor and open-set classifier, combined with a dynamic threshold, are used to identify unknown states in the target domain, completing the diagnosis of bearing open-set data. The experimental results on the CWRU dataset and the self-built motor-bearing platform demonstrate the applicability of WAAN to open-set cross-condition bearing fault diagnosis. These results indicate that the proposed framework has potential for practical condition monitoring and intelligent fault diagnosis of rotating machinery.

Despite these encouraging results, several limitations should be acknowledged. First, the present evaluation focuses mainly on bearing diagnosis under load- and speed-related domain shifts, and the generalizability of WAAN to more complex operating environments requires further validation. Second, although the proposed framework achieves high known-class classification accuracy, its relatively lower unknown-class recall indicates that certain unseen faults remain difficult to detect. Third, the performance of the dynamic threshold depends on the quantile parameter α and sliding-window length W, and a generally applicable label-free parameter selection strategy has not yet been established. In addition, measurement noise and gradually developing faults may alter feature distributions and prediction confidence, potentially reducing the reliability of unknown-fault rejection. The computational inference time and end-to-end detection latency have not been quantitatively evaluated. Although improved fault diagnosis may support condition-based maintenance and operational reliability, energy consumption and conversion efficiency were not directly measured. Therefore, potential energy-efficiency benefits should be regarded as possible practical implications rather than experimentally demonstrated improvements. Future work will focus on improved unknown-fault detection, robust parameter selection, and quantitative evaluation under more challenging operating conditions.

Author Contributions: Conceptualization, X.-M.L. and H.C.; methodology, X.-M.L.; software, X.-M.L.; validation, H.C.; formal analysis, H.C.; writing—original draft, H.C.; writing—review and editing, R.-M.Z. and S.-J.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 52104160, Basic Scientific Research Project of the Liaoning Provincial Department of Education (No. LJ242610147007) and Liaoning Technical University Ordos Research Institute Project for Cultivating University-Local Government S&T Cooperation (YJY-XD-2024-B-017).

Data Availability Statement: The datasets generated and analyzed during the current study are not publicly available because they form part of an ongoing study, but are available from the corresponding author on reasonable request.

Conflicts of Interest: Author Rui-Ming Zhang (who was once a student of Professor Liu Xinming) was employed by State Grid Hebei Electric Power Company, Shijiazhuang, China. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

UDA	Unsupervised domain adaptation
OSDA	Open-Set Domain Adaptation
WAAN	Weighted Adversarial Adaptive Network
1D-CNN	a one-dimensional convolutional neural network
GRL	a gradient reversal layer
CWRU	Case Western Reserve University
t-SNE	t-distributed stochastic neighbor embedding
OSBP	Open Set Domain Adaptation by Backpropagation
DANN	Domain-adversarial training of neural networks
CNN	Convolutional Neural Network

References

1. Chen, X.; Yang, R.; Xue, Y.; Huang, M.; Ferrero, R.; Wang, Z. Deep Transfer Learning for Bearing Fault Diagnosis: A Systematic Review Since 2016. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3508221. [CrossRef]
2. Puntambekar, R.; Vyas, P.; Thakkar, A.; Patel, D. A Survey of Machine Learning and Deep Learning Methods for Vibration-Based Bearing Fault Diagnosis: The Need, Challenges, and Potential Future Research Directions. *Neurocomputing* **2026**, *659*, 131628. [CrossRef]
3. Zhang, W.; Li, X.; Ma, H.; Luo, Z.; Li, X. Open-Set Domain Adaptation in Machinery Fault Diagnostics Using Instance-Level Weighted Adversarial Learning. *IEEE Trans. Ind. Inform.* **2021**, *17*, 7445–7455. [CrossRef]
4. Tian, J.; Han, D.; Karimi, H.R.; Zhang, Y.; Shi, P. Deep Learning-Based Open Set Multi-Source Domain Adaptation with Complementary Transferability Metric for Mechanical Fault Diagnosis. *Neural Netw.* **2023**, *162*, 69–82. [CrossRef] [PubMed]
5. Zhao, J.; Wang, W.; Huang, J.; Ma, X. A Comprehensive Review of Deep Learning-Based Fault Diagnosis Approaches for Rolling Bearings: Advancements and Challenges. *AIP Adv.* **2025**, *15*, 020702. [CrossRef]
6. Tang, H.; Tang, Y.; Su, Y.; Feng, W.; Wang, B.; Chen, P.; Zuo, D. Feature Extraction of Multi-Sensors for Early Bearing Fault Diagnosis Using Deep Learning Based on Minimum Unscented Kalman Filter. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107138. [CrossRef]
7. Siddique, M.F.; Saleem, F.; Umar, M.; Kim, C.H.; Kim, J.-M. A Hybrid Deep Learning Approach for Bearing Fault Diagnosis Using Continuous Wavelet Transform and Attention-Enhanced Spatiotemporal Feature Extraction. *Sensors* **2025**, *25*, 2712. [CrossRef] [PubMed]
8. Wang, G.; Shi, N.; Yang, X.; Liu, Z.; Liu, Z. Bearing Fault Diagnosis Method Based on SAGAN and Improved ResNet. *Sci. Rep.* **2025**, *15*, 33494. [CrossRef] [PubMed]
9. Zhang, W.; Peng, G.; Li, C.; Chen, Y.; Zhang, Z. A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals. *Sensors* **2017**, *17*, 425. [CrossRef] [PubMed]
10. Chen, Z.; Xia, J.; Li, J.; Chen, J.; Huang, R.; Jin, G.; Li, W. Generalized Open-Set Domain Adaptation in Mechanical Fault Diagnosis Using Multiple Metric Weighting Learning Network. *Adv. Eng. Inform.* **2023**, *57*, 102033. [CrossRef]
11. Liu, B.; Li, G.; He, J.; Wang, T.; Shi, R. Complete Knowledge Weighted Adversarial Network and Class Alignment Learning Guided Open Set Multi-Source Domain Adaptive Fault Diagnosis. *Meas. Sci. Technol.* **2025**, *37*, 026005. [CrossRef]
12. Xu, Y.; Chen, L.; Duan, L.; Tsang, I.W.; Luo, J. Open Set Domain Adaptation With Soft Unknown-Class Rejection. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 1601–1612. [CrossRef] [PubMed]
13. Lei, T.; Pan, F.; Hu, J.; He, X.; Li, B. A Fault Diagnosis Method for Rolling Bearings in Open-Set Domain Adaptation with Adversarial Learning. *Sci. Rep.* **2025**, *15*, 10793. [CrossRef] [PubMed]
14. De Maesschalck, R.; Jouan-Rimbaud, D.; Massart, D.L. The Mahalanobis Distance. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 1–18. [CrossRef]
15. Oh, H.; Lee, Y.; Lee, J.; Joo, C.; Lee, C. Feature Selection Algorithm Based on Density and Distance for Fault Diagnosis Applied to a Roll-to-Roll Manufacturing System. *J. Comput. Des. Eng.* **2022**, *9*, 805–825. [CrossRef]
16. Jiang, M.; Xiang, Y.; Sheng, C. Based on Spatial Mahalanobis Distance: A Novel Zero-Shot Learning Method for Compound Fault Identification and Decoupling. *Expert Syst. Appl.* **2025**, *272*, 126680. [CrossRef]
17. Guo, Y.; Zhang, J.; Sun, B.; Wang, Y. Adversarial Deep Transfer Learning in Fault Diagnosis: Progress, Challenges, and Future Prospects. *Sensors* **2023**, *23*, 7263. [CrossRef] [PubMed]

18. He, Z.; Shao, H.; Zhong, X.; Zhao, X. Ensemble Transfer CNNs Driven by Multi-Channel Signals for Fault Diagnosis of Rotating Machinery Cross Working Conditions. *Knowl.-Based Syst.* **2020**, *207*, 106396. [CrossRef]
19. Wen, L.; Li, X.; Gao, L.; Zhang, Y. A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Trans. Ind. Electron.* **2018**, *65*, 5990–5998. [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.