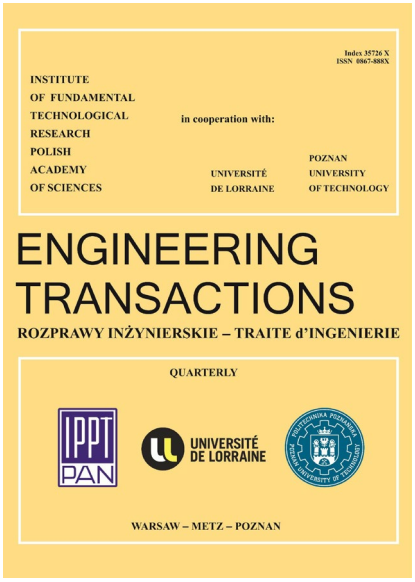
JOURNAL PRE-PROOF

This is an early version of the article, published prior to copyediting, typesetting, and editorial correction. The manuscript has been accepted for publication and is now available online to ensure early dissemination, author visibility, and citation tracking prior to the formal issue publication.

It has not undergone final language verification, formatting, or technical editing by the journal's editorial team. Content is subject to change in the final Version of Record.

To differentiate this version, it is marked as "PRE-PROOF PUBLICATION" and should be cited with the provided DOI. A visible watermark on each page indicates its preliminary status.

The final version will appear in a regular issue of *Engineering Transactions*, with final metadata, layout, and pagination.



Title: A Small-Sample Fault Diagnosis Method for Rolling Bearings Based on Balanced Distribution Adaptation and Support Vector Machine

Author(s): Dong Chen, Mingshan Zhang, Yaguang Gao

DOI: https://doi.org/10.24423/engtrans.2026.3649

Journal: *Engineering Transactions*

ISSN: 0867-888X, e-ISSN: 2450-8071

Publication status: In press

Received: 2025-08-31

Accepted: 2025-12-02

Published pre-proof: 2026-04-07

Please cite this article as:

Chen D., Zhang M., Gao Y., A Small-Sample Fault Diagnosis Method for Rolling Bearings Based on Balanced Distribution Adaptation and Support Vector Machine, *Engineering Transactions*, 2026, https://doi.org/10.24423/engtrans.2026.3649

A Small-Sample Fault Diagnosis Method for Rolling Bearings Based on Balanced Distribution Adaptation and Support Vector Machine

Dong Chen*, Mingshan Zhang, Yaguang Gao

Jinan Technician College, Handan, 056000, China

*Corresponding email: chendong859052@163.com

Abstract: To address the issue of low diagnostic accuracy caused by distribution differences between source and target domains in rolling bearing fault diagnosis, this study proposes a method combining Balanced Distribution Adaptation (BDA) and Support Vector Machines (SVM). The approach utilizes BDA to simultaneously minimize discrepancies in both edge and conditional distributions between domains, enabling effective feature alignment and enhancing the model's cross-domain generalization in small-sample scenarios. After extracting time and frequency-domain features, BDA adaptively adjusts feature distributions, and SVM is employed for fault classification. Experimental results demonstrate that the BDA-SVM method achieves over 94% diagnostic accuracy, showcasing strong performance and robustness for bearing fault diagnosis. Compared with traditional SVM and other methods without transfer learning, the proposed approach shows significant improvement in diagnostic accuracy under cross-domain conditions.

Keywords: rolling bearings; fault diagnosis; balanced distribution adaptation; transfer learning; support vector machines

Introduction

Bearings in industrial equipment generally play a critical role in support and transmission, and their health status directly affects the operational stability and service life of the entire equipment [1-2]Especially in critical fields such as mechanical manufacturing, wind power, and rail transportation, sudden failures of rolling bearings can lead to equipment shutdowns, causing significant economic losses and even safety accidents. However, due to their long-term operation in high-temperature, heavy-load, and high-frequency alternating load environments, rolling bearings are prone to various

faults [2-3]. Existing research indicates that approximately 30% of failures in rotating machinery are caused by rolling bearing failures [1, 4]. Therefore, achieving precise early fault diagnosis[21] is of great significance for ensuring normal equipment operation, extending equipment service life, and preventing loss expansion [5].

Traditional fault diagnosis relies on manual experience analysis, which is inefficient and highly subjective, and requires precise mechanistic models as support, presenting numerous challenges in practical applications [6]. With technological advancements and accumulated experience, data-driven intelligent diagnostic methods, leveraging machine learning and deep learning's ability to fit complex nonlinear functions, have demonstrated outstanding performance in industrial equipment fault diagnosis [4].

However, the performance of data-driven methods highly depends on large-scale labeled data, which is a significant bottleneck in industrial settings due to the scarcity of fault data. While normal operation data is relatively abundant in most industrial scenarios, fault data is often scarce due to various challenges such as difficulty in detection, characterization, and reproduction [6]. Existing supervised learning models rely on large amounts of labeled data and are prone to overfitting under small-sample conditions, significantly reducing diagnostic accuracy.

Besides, recent studies have further explored transfer learning for fault diagnosis. For instance, a comparative study [7] is conducted on classical shallow transfer learning methods using the CWRU dataset, including BDA+KNN and BDA+SVM, and reported that BDA+KNN achieved the highest accuracy while BDA+SVM ranked second. Similarly, Lei et al. [8] provides a comprehensive review of deep-learning-based fault diagnosis methods and highlights their ability in feature extraction under varying working conditions. While these deep learning-based methods have shown promising performance, they often require a larger number of parameters and computational resources, and their performance can degrade in extreme small-sample scenarios.

To address the issue of insufficient model generalization under small sample conditions, transfer

learning provides a solution. Its core lies in reducing the distribution differences between the source domain and the target domain to achieve knowledge transfer, enabling the model to effectively reuse source domain knowledge in the target domain [7]. Existing studies mostly focus on the marginal distribution features between the source domain and the target domain, neglecting the conditional distribution features [1,8]. Balanced Distribution Adaptation (BDA) achieves effective cross-domain feature transfer by jointly optimizing the differences between edge distributions and conditional distributions, and introducing a tuning parameter to balance the weights of edge distributions and conditional distributions. This approach preserves category discrimination information while minimizing domain-specific bias, providing an effective method for fault diagnosis across different operating conditions and devices [8,9-10].

In specific fault diagnosis tasks, the model's ability to fit nonlinear functions is key. This ability is usually achieved through deep learning, random forests, and support vector machines (SVM) [11-13]. Among them, SVM is a classic supervised learning method with flexible kernel functions and penalty factors, which has long been stable in binary and multi-class classification problems [15]. Its core idea is to maximize the separation between different categories of samples by constructing an optimal hyperplane to achieve classification [16].

Given this, this study proposes a rolling bearing fault diagnosis method combining BDA and SVM: the BDA algorithm is used to jointly align the edge distributions and conditional distributions of the source domain and target domain, addressing the limitation of existing studies that only consider edge distributions to achieve feature layer alignment; simultaneously, SVM is utilized to leverage its advantages in classification tasks. The synergistic effect of the two methods not only alleviates the data scarcity constraint through transfer learning but also ensures diagnostic accuracy by leveraging the strong classification capabilities of SVM, ultimately achieving high-precision fault diagnosis under cross-condition and small-sample conditions.

1 Signal Time-Frequency Domain Feature Analysis and Extraction

The essence of SVM classification is eigenvector-oriented, while the extracted raw signals are discrete time series. Therefore, time-domain and frequency-domain feature extraction (such as mean value, root mean square, spectral energy, etc) is required to convert them into quantitative indicators that reflect the key features of the signals. These indicators are first aligned between the source domain and target domain using BDA, then used as the training basis for SVM.

Time-domain signals and frequency-domain signals are linearly correlated through the Fourier transform, reflecting signal characteristics from different dimensions. This paper intends to integrate multiple time-domain and frequency-domain features as the inputs of BDA [22]and SVM.

1.1 Time-domain feature extraction

Time-domain signals are used to describe the amplitude, trend, and changes of a signal. They can intuitively reflect the original characteristics of the signal and are commonly used to describe the overall level and sudden faults of vibration signals [17]. Time-domain features are directly extracted from the time series of vibration signals. Feature indicators such as the mean, root mean square, and variance are commonly used in signal analysis and also have practical significance in engineering.

The mean value is used to reflect the overall level of a signal. In bearing rotor systems, it can indirectly indicate bearing misalignment, etc. Its definition is:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (1)$$

In the formula, x_i is the sample value of the vibration signal, and N is the total number of sampling points in the signal.

The root mean square (RMS) value is used to measure the energy level of a signal. A higher RMS value typically indicates that the equipment is experiencing significant vibration. Its definition is:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \qquad (2)$$

Skewness (S) is a statistical measure describing the symmetry of a signal, used to analyze signal

asymmetry. Its expression is:

$$S = \frac{1}{N}\sum_{i=1}^{N}(\frac{x_i - \mu}{\sigma})^3 \qquad (3)$$

In the formula, σ is the standard deviation of the signal.

Kurtosis (K) is a statistical measure of signal sharpness, indicating whether the signal contains sudden anomalies or impact faults. Its expression is:

$$K = \frac{1}{N}\sum_{i=1}^{N}(\frac{x_i - \mu}{\sigma})^4 \qquad (4)$$

Variance (V) is used to describe the degree of fluctuation in a signal. The larger the variance, the more intense the vibration and the higher the likelihood of a fault. Its expression is:

$$V = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \qquad (5)$$

The crest factor (C) is defined as the ratio of the maximum value of the signal to its root mean square value. It is typically related to the impact the system experiences and can be used to assess the intensity of transient impacts in the signal. Its expression is:

$$C = \frac{\max(|x_i|)}{X_{RMS}} \qquad (6)$$

1.2 Frequency Domain Feature Extraction

Frequency domain features can effectively capture the periodicity or characteristic frequency information of equipment faults [18] and are suitable for detecting fault modes in specific frequency bands. Various faults in rolling bearings have typical frequency characteristics and are closely related to the inner and outer rings, rolling element dimensions, contact methods, and relative rotational frequency [19]. This paper extracts two common frequency domain features: spectrum center frequency and spectrum energy.

Spectral center frequency (Center Frequency, CF) is used to describe the central position of the frequency domain signal, i.e., the average value of the signal frequency distribution, and can indicate

the dominant frequency where energy is concentrated. Its expression is:

$$CF = \frac{\sum_{i=1}^{N} f_i P(f_i)}{\sum_{i=1}^{N} P(f_i)} \qquad (7)$$

In the formula,*fi* is the frequency component of the signal, and P(*fi*) is the corresponding power spectral density.

Faults cause significant enhancement of energy at the feature frequency location [19]. Spectral energy (SE) represents the total energy of a signal in the frequency domain and is commonly used to assess signal intensity across different frequency bands. Its expression is:

$$SE = \sum_{i=1}^{N} \left| X(f_i) \right|^2 \qquad (8)$$

In the formula, X(*fi*) is the amplitude of the signal's Fourier transform spectrum.

2 BDA-SVM Algorithm Principle

2.1 BDA Transfer Learning Principle and Framework

Traditional machine learning methods typically assume that the data distributions of the training set and the test set are consistent. However, in practical applications, there are inevitably significant differences in the distributions of the source domain and the target domain [7-9] . The goal of transfer learning is to mitigate the problem of insufficient samples in the target domain by acquiring knowledge from the source domain and transferring it to the target domain, thereby improving learning performance. Its core idea is to utilize knowledge learned in the source domain (such as model parameters, feature distributions, or implicit space representations) to improve the performance of tasks in the target domain, enabling models trained on the source domain to be effectively applied in the target domain after fine-tuning.

Let the source domain and target domain datasets be denoted as:

$$D_s = \{(x_s^{(i)}, y_s^{(i)})\}_{i=1}^{n_s}, \qquad (9)$$

$$D_t = \{x_t^{(j)}\}_{j=1}^{n_t}, \quad (10)$$

In the formula, x_i^s and x_i^t represent the sample features in the source domain and target domain datasets, respectively, n_s and n_t are the number of samples in the source domain and target domain, respectively, and y_i^s are the corresponding sample labels. In practical applications, the labels y_j^T in the target domain are often unknown. We employ pseudo-labeling techniques to estimate the conditional distribution, where initial predictions from the source domain model are used to approximate the true labels.

The feature distributions of the source domain and target domain, $P_S(X)$ and $P_T(X)$, as well as the joint distributions, $P_S(X,Y)$ and $P_T(X,Y)$, are typically different. Therefore, the goal of transfer learning is to train a model, $f_S(x)$, in the source domain that can perform well in the target domain, i.e., $f_T(x) \approx f_S(x)$. Here, $f_T(x)$ represents the model performance in the target domain, which is the ultimate objective of transfer learning.

2.2 Introduction to BDA Algorithms

BDA is a domain adaptation technique that simultaneously minimizes the differences between marginal distributions and conditional distributions, thereby aligning the source and target domains in terms of feature dimensions and enhancing the transferability of cross-domain data [1,8,20]. The original BDA method was proposed by Wang et al. [25], who introduced a framework for simultaneously minimizing marginal and conditional distribution discrepancies. Its objective is to find a mapping function $\Phi(X)$ such that the source domain D_S and the target domain D_T have more similar distributions in the mapping space.

This is achieved by simultaneously optimizing two objectives: minimizing the Marginal Distribution Discrepancy (MDD) and the Conditional Distribution Discrepancy (CDD). The MDD reflects the distribution deviation of data features between the source and target domains across the

entire space, while the CDD considers the distribution differences between data of the same category in the source and target domains, ensuring that classification information remains consistent during the transfer process. The expressions are as follows:

$$MDD = \left\| \mathrm{E}_{x_i^S \sim P_S(X)} \left[\Phi(x_i^S) \right] - \mathrm{E}_{x_i^T \sim P_T(X)} \left[\Phi(x_i^T) \right] \right\|^2 \tag{11}$$

$$CDD = \sum_{c \in C} \left\| \mathrm{E}_{x_i^S \sim P_S(X|Y=c)} \left[\Phi(x_i^S) \right] - \mathrm{E}_{x_i^T \sim P_T(X|Y=c)} \left[\Phi(x_i^T) \right] \right\|^2 \tag{12}$$

In practical applications where the target domain labels are unknown, BDA employs an iterative pseudo-labeling strategy. Initially, a base classifier (e.g., SVM) trained on the source domain is used to predict pseudo-labels for the target domain samples. These pseudo-labels are then used to estimate the conditional distribution $P(y^T \mid x^T)$. It is important to note that while a constant mapping function ϕ might trivially minimize distribution discrepancies, BDA avoids this degenerate solution by incorporating constraints that preserve the intrinsic data structure and variance. The optimization framework ensures that the mapping ϕ not only aligns the distributions but also maintains the discriminative information necessary for classification. During the BDA optimization process, the pseudo-labels are iteratively refined as the feature alignment improves, leading to more accurate conditional distribution matching. This approach allows BDA to effectively handle the conditional distribution adaptation even when target domain labels are unavailable during training.

BDA introduces a balancing factor λ to control the balance between MDD and CDD:

$$L_{BDA} = MDD + \lambda CDD \tag{13}$$

In the formula, λ is a hyperparameter used to adjust the influence of marginal distribution and conditional distribution on the transfer. By adjusting λ, the model's adaptability to distribution differences between the source domain and target domain can be controlled. The core idea of BDA is to simultaneously coordinate the two objective parameters MDD and CDD by introducing the parameter λ.

2.3 SVM Model Construction

SVM is a classic supervised learning method commonly used for binary classification problems. The SVM classifier was originally introduced by Cortes and Vapnik [26], and the kernel trick for nonlinear mapping follows the formulation described by Schölkopf and Smola [27]. Its core idea is to maximize the distance between samples of different categories by constructing a hyperplane, thereby achieving classification. The basic objective is to find an optimal hyperplane that separates the different categories in the dataset [13]. Although SVM is fundamentally a binary classifier, it can be extended to multi-class problems using strategies such as one-vs-rest or one-vs-one. In this study, we employ the one-vs-rest approach for the four-class bearing fault diagnosis task. Assume that the data points in the dataset are (x_i, y_i) , where $x_i \in R^n$ are the input feature vectors, and y_i are the corresponding class labels. The SVM classification hyperplane can be represented as:

$$d = \frac{\left| w^T x_i + b \right|}{\left\| w \right\|} \qquad (14)$$

In the formula, w is the normal vector of the hyperplane, and b is the offset.

For each sample point x_i , the distance to the hyperplane is:

$$\frac{y_i (w^T x + b)}{\left\| w \right\|}. \qquad (15)$$

To maximize this margin while ensuring correct classification, the concept of soft margin is introduced to handle points on or inside the boundary, which determines the final position of the classification hyperplane. To find the optimal classification hyperplane, SVM solves the following optimization problem:

$$\min_{w,b} \frac{1}{2} \left\| w \right\|^2, \qquad (16)$$

under the constraints:

$$y_i (w^T x_i + b) \geq 1, i = 1, 2, ..., n \qquad (17)$$

In the equation, $\frac{1}{2} \left\| w \right\|^2$ is the regularization term, which prevents overfitting and ensures that the

classifier has good generalization ability.

However, in practical applications, many data points cannot be well separated by a single linear hyperplane. To address this issue, SVM introduces kernel functions, which map input data to a higher-dimensional feature space, thereby achieving linear separability in high dimensions. Since the RBF kernel function effectively handles nonlinear classification problems and exhibits strong classification capabilities in high dimensions, it is selected as the built-in kernel function for the SVM model in this paper, with the following expression:

$$K(x_i, x_j) = \exp(-\gamma \left\| x_i - x_j \right\|^2) \quad (18)$$

Typically, in SVM, a slack variable ξ_i is introduced to find the optimal hyperplane. After introduction, the optimization problem can be expanded as:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i , \quad (19)$$

In the formula, C is the penalty parameter, which controls the trade-off between margin and classification error. By adjusting C , a balance can be found between maximizing the margin and minimizing the classification error. Therefore, the final decision function of the SVM can be expressed as:

$$f(z) = sign(\sum_{i=1}^{n} \alpha_i y_i K(z_i, z) + b) \quad (20)$$

In the formula, α_i is the Lagrange multiplier, obtained by solving the optimization problem.

2.4 BDA-SVM Rolling Bearing Fault Diagnosis Strategy

Combining the advantages of BDA and SVM, this study proposes a BDA-SVM-based rolling bearing fault diagnosis strategy to address the fault diagnosis problem when there is significant feature deviation between the source domain and the target domain. The model flow is shown in Figure 1. The white boxes represent the output and input, while the black boxes represent the intermediate processes.

As shown in Figure 1, first, vibration signals from open-source data and rolling fault vibration signals are acquired. Time-domain feature extraction and frequency-domain feature extraction are then performed on these two types of signals. The time-domain features used in this example typically include mean, variance, skewness, and kurtosis, while the frequency-domain features involve frequency components and power spectral density information. The extracted feature signals are divided into source domain feature samples and target domain feature samples, corresponding to the open-source data and rolling bearing fault data, respectively. Then, the BDA algorithm is applied to adjust and transfer the feature distributions of the source domain and target domain. Finally, the feature-transferred data is input into the SVM classifier to train the fault classification and diagnosis model. The SVM model is then trained and classified based on the input target domain features to identify different fault types and ultimately complete the fault diagnosis task.

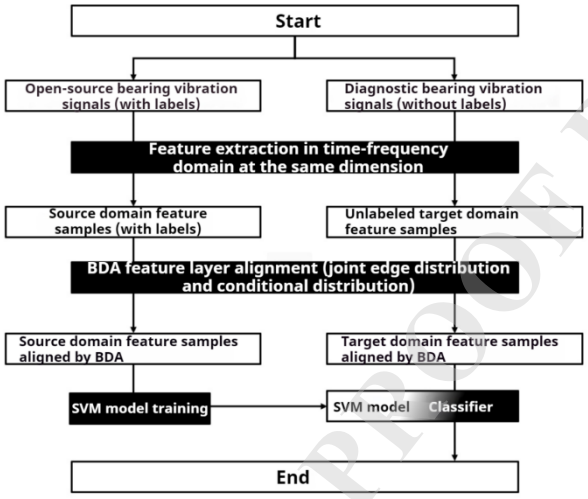


Fig. 1 BDA-SVM fault diagnosis flow chart

3 Experiments and Analysis

3.1 Experimental Setup

The experiment uses the MATLAB platform for algorithm development, data processing and visualization operations. Machine learning models were trained and tested using MATLAB toolboxes, while signal processing tools were employed for feature extraction to ensure accurate analysis and processing of experimental data. The vibration signals were sampled at 12 kHz with a duration of 1

second per sample, resulting in 12,000 data points per signal. Sequential sampling was used without overlap between segments. To validate the effectiveness of the proposed method in rolling bearing fault diagnosis, the Case Western Reserve University (CWRU) dataset was used. The CWRU dataset includes four main fault modes: normal, inner ring fault, outer ring fault, and rolling element fault. The fault category codes are shown in Table 1.

Tab. 1 Bearing sample set fault number

Fault type	Fault Number
Normal	0
Outer ring	1
Inner ring	2
Roller	3

Inner ring failures manifest as physical damage to the inner ring of the bearing, while outer ring failures are associated with damage to the outer ring. Rolling element failures indicate abnormal damage to the balls or rollers. Each failure mode includes different damage levels (such as 0.007 inches, 0.014 inches, 0.021 inches), and data is collected by recording vibration signals during bearing operation using an accelerometer.

Additionally, the experiment utilized rolling fault data collected from actual industrial equipment, which served as the target domain to validate the model's transferability in real-world industrial applications. The experimental setup is shown in Figure 2, employed SKF 6205 deep groove ball bearings. Faults were artificially introduced using electro-discharge machining to create single-point defects on the inner race, outer race, and rolling elements, with fault diameters of 0.007, 0.014, and 0.021 inches respectively. Vibration acceleration signals were acquired using a piezoelectric accelerometer (model: PCB 352C33) mounted vertically on the drive-end bearing housing. The signals were sampled at 12 kHz using a 16-bit data acquisition system. The experimental setup is shown in Figure 2.

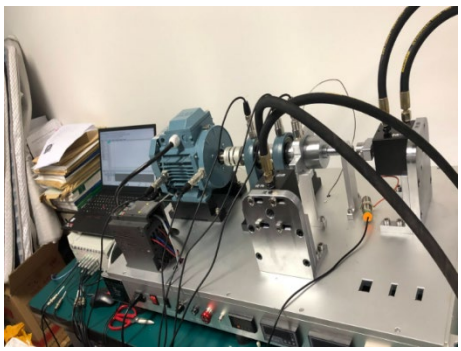
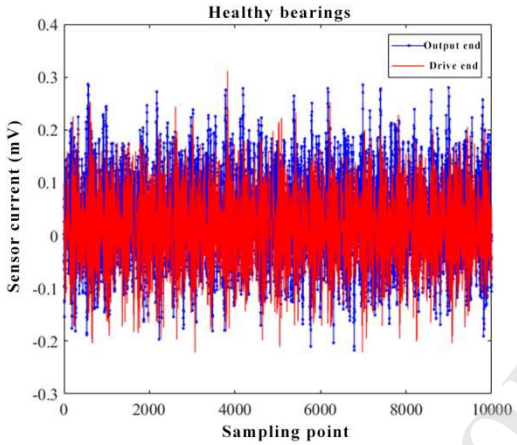


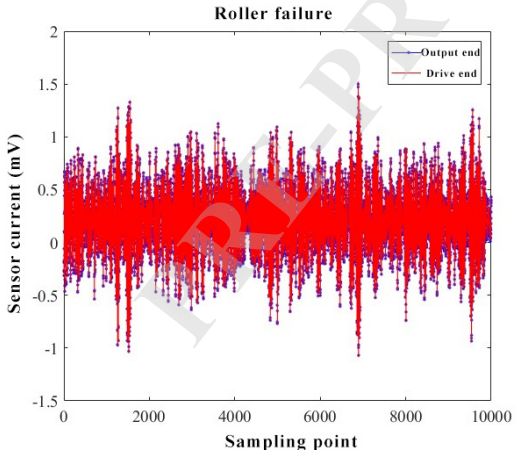
Fig. 2 Bearing signal acquisition device

3.2 Time-frequency domain responses of normal and faulty bearings

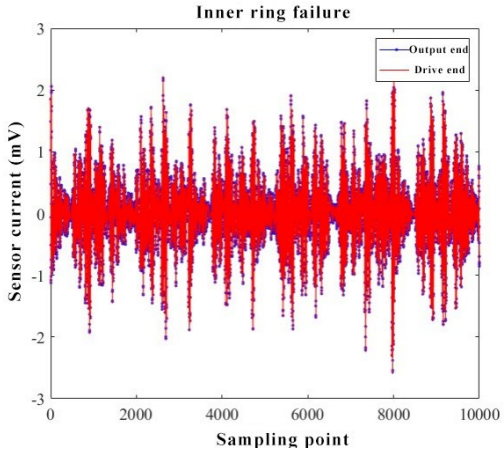
The signals acquired by the device are shown in Figure 3.



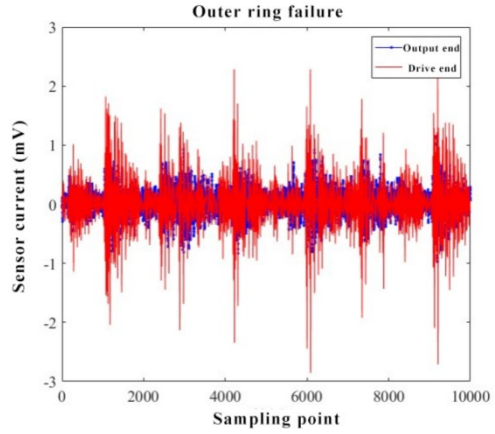
(a) Normal signal



(b) Rolling element fault



(c) Inner ring fault



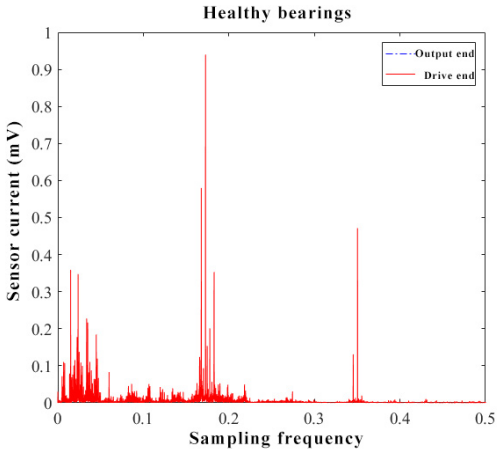
(d) Outer ring fault

Fig. 3 Signal diagram collected by experiment

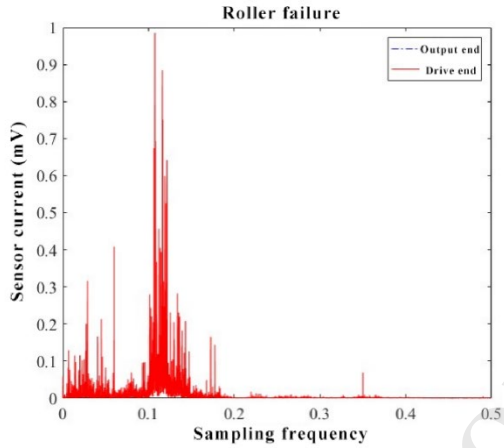
Figure 3 displays representative time-domain signals from both the CWRU dataset and the in-house collected industrial data to illustrate characteristic waveform patterns: (a) normal condition, (b) rolling element fault, and (c) inner ring fault are from the CWRU dataset, while (d) outer ring fault is from the industrial data collected in-house. The signals acquired by the device are shown in Figure 3. By comparing the different faults in Figure 3, it can be observed that any fault causes an increase in the amplitude of the response in the time domain. Under normal conditions, the maximum amplitude is approximately 0.2 mV, while under fault conditions, the maximum amplitude increases by approximately 5 to 40 times, and the low-frequency harmonic components are significantly enhanced, indicating the rationality of using time-frequency domain features.

To investigate the frequency components of the signal, we normalized the frequency domain

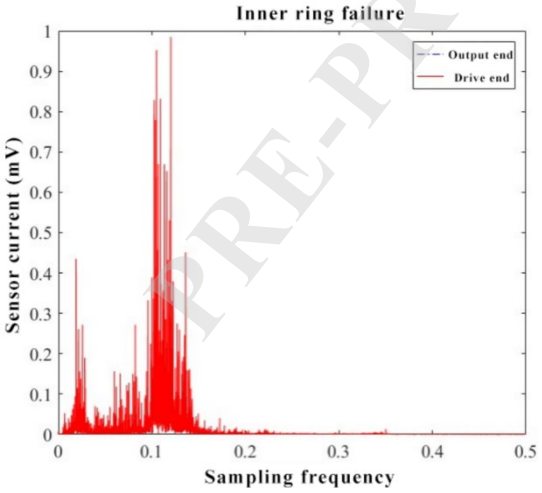
features, and the resulting frequency domain signal is shown in Fig. 4.



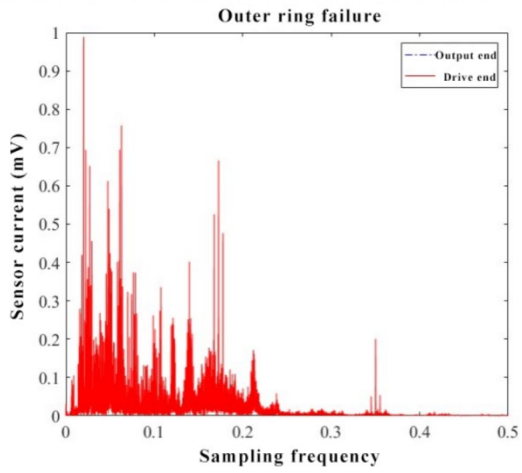
(a) Normal signal



(b) Rolling element fault



(c) Inner ring fault



(d) Outer ring fault

Fig. 4 Frequency domain features of signals acquired in the experiment

As shown in Figure 4(a), the characteristic frequencies of a healthy bearing are clearly visible, with the fundamental frequency (0.17) and overtone (0.35) being prominent, exhibiting a typical frequency modulation phenomenon. This is related to the time-varying support stiffness characteristics of rolling bearings, and the wide-frequency characteristics are not prominent. Under rolling element failure conditions, the amplitude near the characteristic frequency (0.12) of the rolling elements is significantly higher than the fundamental frequency (0.17). Under inner ring failure conditions, the energy is primarily concentrated in the low-frequency range, and the fundamental frequency cannot be identified from the frequency domain signals. Under outer ring fault conditions, the wide-frequency characteristics are more pronounced, and the frequency components in the response are more complex. Although the fundamental frequency and its harmonics can be identified, the formation mechanisms of other frequency components cannot be inferred. This is because the outer ring is closer to the sensor, resulting in a more pronounced response to fault-induced impacts. As shown in Figure 3(d), the fault signals are easier to identify in the time-domain signals compared to Figure 4(d) in the frequency-domain signals. Therefore, using time-frequency domain characteristics as diagnostic criteria is more reasonable.

3.3 Validation of the BDA Method

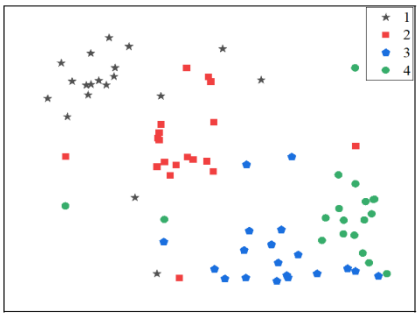
In this experiment, the source domain consists of the CWRU dataset with specific operating

conditions, while the target domain comprises both different operating conditions from CWRU and industrial field data, creating a challenging cross-domain scenario. The dataset composition and division are as follows. The source domain (CWRU dataset) comprises 800 samples, with 200 samples for each health state (normal, inner ring fault, outer ring fault, and roller fault). The target domain (industrial field data) contains 400 samples, with 100 samples per category. To ensure the representativeness of the data and to maintain the class distribution in both domains, we employed a stratified random sampling strategy to divide the datasets. To achieve effective transfer, 70% of the source domain data (560 samples) was used for training and 30% (240 samples) for validation. The target domain data was similarly divided into a 70% training set (280 samples) and a 30% test set (120 samples) to evaluate the model's performance in cross-domain fault diagnosis. It is important to note that during the BDA adaptation process, the labels of the target domain training set are treated as unknown and are estimated using pseudo-labels generated by the source domain model. The final diagnostic accuracy is reported based on the predictions on the target domain test set using the ground-truth labels.

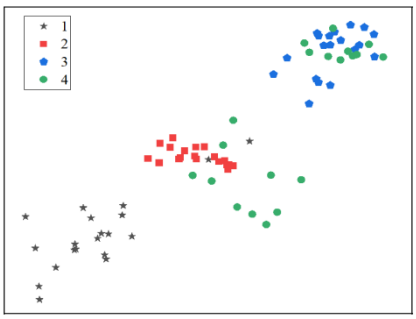
In the BDA algorithm parameter settings, the regularization parameter λ is used to balance the distribution differences and the relationship between the feature matrices. Its value typically ranges from 0.1 to 1 and is optimized through experiments to determine the optimal value. The embedding dimension d determines the dimension of the subspace after feature mapping. A larger dimension retains more feature information but increases computational complexity. The optimal dimension is determined through cross-validation in the experiments. The number of iterations T in BDA is set to 100 to ensure the minimization of feature distribution differences.

In actual bearing fault diagnosis, the source domain and target domain data often exhibit significant distribution differences, which can lead to poor performance of traditional models when applied across domains. Therefore, this paper further optimizes feature transfer effects by adaptively adjusting the feature distributions of the source domain and target domain, ensuring better classification performance in fault diagnosis in the target domain.

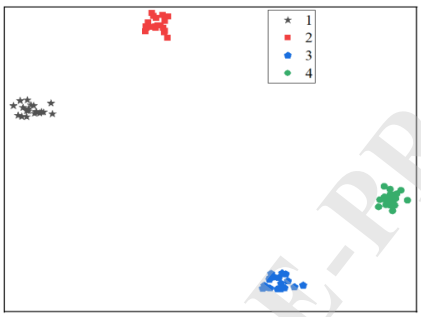
This paper uses the Transfer Component Analysis (TCA) transfer learning method as an example to compare the performance improvement of the proposed method. Figure 4 shows the visualization results of BDA and TCA before and after feature transfer.



(a) Before feature transfer



(b) After TCA feature transfer



(c) After BDA feature transfer

Fig. 5 Visualization before and after transfer

As can be clearly seen from the classification results in Fig. 5(a), before feature migration, due to the significant differences in data distribution between the source domain and the target domain, data points of different categories are mixed together in the feature space, resulting in blurred classification boundaries and making it difficult to perform effective fault classification. After

applying the TCA method to migrate features (Fig. 5(b)), the distribution of sample data from different fault categories in the feature space improved, with some categories beginning to cluster together and distribute more densely. This is because TCA can reduce the distribution differences between the source domain and target domain. However, some categories still overlap in the figure, indicating that TCA has limited effectiveness in adjusting cross-domain feature distributions. This is because TCA does not adequately consider the imbalance in feature distributions between the source domain and the target domain, resulting in insufficient separation of some data categories and thereby affecting classification accuracy.

After feature transfer using the BDA method (Figure 5(c)), the data distribution of the source domain and target domain was significantly optimized. Data points of different fault categories became more compact and clear in the feature space, and the distribution differences between the source domain and target domain were greatly reduced. BDA achieves this by adaptively adjusting the feature distributions between the source and target domains, making samples of the same category more consistently distributed across domains and enhancing the boundaries between categories. This significantly improves classification accuracy and stability. This indicates that BDA is indeed effective in cross-domain fault diagnosis, effectively addressing distribution differences and aiding in resolving the small-sample problem in engineering applications. Figure 5 visualizes the feature distribution of the source and target domain data. The original feature space consists of the 9 time-frequency domain features extracted in Section 1 (including mean, root mean square, skewness, kurtosis, variance, crest factor, spectral center frequency, and spectral energy). To visualize these high-dimensional features in a 2D plane, we employed the t-SNE (t-distributed Stochastic Neighbor Embedding) dimensionality reduction method, which is widely used for visualizing high-dimensional data while preserving the structure of local neighborhoods.

3.4 Performance Validation of BDA-SVM

To validate the performance of the proposed BDA-SVM model, this paper designed a series of comparative experiments to compare the diagnostic effectiveness of multiple traditional classification

algorithms, including decision trees (DT), k-nearest neighbors (KNN), and the original SVM. The purpose of the experiment is to demonstrate the effectiveness of BDA-SVM in handling distribution inconsistency issues through model comparisons and to evaluate its diagnostic accuracy in actual fault classification tasks. The experimental results are shown in Figure 6.

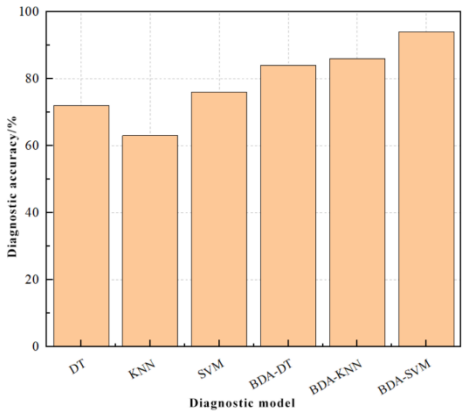


Fig. 6 Diagnostic results of different models

As can be seen from the comparison results in Figure 6, the traditional DT and KNN models perform the worst in cross-domain fault diagnosis tasks. Specifically, the diagnostic accuracy of KNN is only 63%, while the accuracy of the DT model is slightly higher at 72%. The SVM achieves the highest diagnostic accuracy, but it still only reaches 76%. These traditional models cannot effectively adapt to the target domain data without addressing data distribution differences, resulting in poor cross-domain diagnostic performance. These results indicate that relying solely on traditional machine learning classifiers is insufficient to address the distribution inconsistency issue in cross-domain fault diagnosis, resulting in limited classification performance. As shown in the figure, incorporating BDA into traditional models significantly improves diagnostic accuracy, fully demonstrating the feasibility of combining BDV with classical machine learning and further validating the conclusions drawn in the previous section. For the multi-class problem, we employed one-vs-rest SVM with RBF kernel. The results presented are based on the CWRU dataset under cross-operating-condition scenarios. Detailed confusion matrices for each method are available in the supplementary materials.

In contrast, the proposed BDA-SVM model demonstrates outstanding advantages in cross-

domain fault diagnosis tasks. By introducing the BDA algorithm before the SVM model to adaptively adjust the feature distributions of the source and target domains, the diagnostic accuracy of the BDA-SVM model is significantly improved, increasing from 76% when using SVM alone to over 94%. This is because the distribution differences between the source domain and the target domain are significantly reduced after BDA processing, enabling the SVM to better adapt to the target domain data. Compared with the comparison models, whether it is DT, KNN, or the original SVM, BDA-SVM performs better in cross-domain fault diagnosis. Not only is the accuracy greatly improved, but the distribution robustness is also inevitably enhanced. This result indicates that the BDA-SVM model has stronger comprehensive performance in practical applications.

It is important to note that the performance ranking between BDA combined with SVM and BDA combined with KNN in our experiments differs from that reported in earlier comparative studies on laboratory datasets. In those studies[7], BDA with KNN achieved the highest accuracy because the source and target data were collected entirely under controlled laboratory conditions. The transferred features formed compact and well-separated clusters, enabling KNN to fully leverage its neighborhood-based decision mechanism.

In our work, however, the target domain includes real industrial data with stronger noise, more complex operating conditions, and less consistent local feature density after domain adaptation. Under such circumstances, KNN becomes more sensitive to irregular neighborhood structures and boundary noise, which reduces its classification stability. In contrast, SVM relies on margin maximization and is less affected by local density fluctuations, resulting in higher robustness and better performance in Fig. 6.

4 Conclusion

This paper addresses the distribution differences between source and target domains in cross-domain fault diagnosis and proposes a rolling bearing fault diagnosis strategy combining BDA-SVM. The BDA algorithm effectively reduces the distribution differences between the source and target

domains at the feature level, thereby enhancing the model's generalization ability in cross-domain applications. The experimental results show that compared with traditional machine learning algorithms (such as DT, KNN, SVM, etc.) that do not introduce transfer learning, BDA-SVM has significant advantages in cross-domain fault diagnosis tasks, with a significantly improved diagnosis accuracy of over 94%, demonstrating strong robustness and generalization capabilities.

Future research can build upon the existing model to further explore how to enhance the effectiveness of the method and apply it to more industrial scenarios. Combining deep learning techniques [23] or other advanced transfer learning methods [24] can further improve the model's performance in more complex and diverse cross-domain fault diagnosis tasks. However, this study has certain limitations. The method's performance under extreme distribution shifts requires further investigation. Future work will focus on combining BDA-SVM with deep feature extraction techniques and extending the approach to more complex industrial scenarios with multiple fault modes and severe data imbalance.

Acknowledgements/ Funding

None.

Declaration of Conflicting Interest

The Author(s) declare (s) that there are no known competing financial interests or personal relationships that could influence the work reported in this paper.

References

1 Xu H, Zhao Z., Xiao X., Wang Z., Bearing Fault Diagnosis Method Based on Multi-Adversarial and Balanced Distribution Adaptation [in Chinese], *Journal of Vibration and Shock*, **44**(05): 302-313, 2025.

2 Lu H., Zheng D., Liu Y., Chang B., Bearing Fault Diagnosis Method Based on Multi-component Sparse Representation [in Chinese], *Journal of Harbin University of Commerce (Natural Science Edition)*, **41**(04): 442-449, 2025, https://doi.org/10.19492/j.cnki.1672-0946.2025.04.012

3 Chen J., Tang Z., Zhou J., Rolling Bearing Fault Diagnosis Based on Improved SDP and FasterNet-GCAM [in Chinese], *Modern Manufacturing Engineering*, (07): 129-138+41, 2025, https://doi.org/10.16731/j.cnki.1671-3133.2025.07.016

4 Zhang Y., Research Progress on Rolling Bearing Fault Diagnosis Based on 2D Data Sets [in Chinese], *Mechanical Management and Development*, **40**(07): 114-117+320, 2025, https://doi.org/10.16525/j.cnki.cn14-1134/th.2025.07.042

5 Chen R., Zhu J., Hu X., Wu H., Fault diagnosis method of rolling bearing based on multiple classifier ensemble of the weighted and balanced distribution adaptation under limited sample imbalance, *ISA Transactions*, (114):434–443, 2021. https://doi.org/10.1016/j.isatra.2020.12.034

6 Zhou Q., Ma W., Zhang Y., Guo J., A multi-source domain adaptation approach with learning domain-specific representations for bearing fault diagnosis under limited samples, *Applied Soft Computing Journal*, (184): 113727, 2025, https://doi.org/10.1016/j.asoc.2025.113727

7 Chen R., Zhu J., Hu X., Wu H., Fault diagnosis of rolling bearing based on multiple classifier ensemble of the weighted and

balanced distribution adaptation under limited sample imbalance, *ISA Transactions*, 114: 434 – 443, 2021, https://doi.org/10.1016/j.isatra.2020.12.034
8 Lei Y., Yang B., Jiang X., Jia F., Li N., Nandi A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap, *Mechanical Systems and Signal Processing*, 138: 106587, 2020. https://doi.org/10.1016/j.ymssp.2019.106587
9 Chen R., Zhu Y., Hu X., Zhao S., Zhang X., Rolling Bearing Fault Diagnosis under Different Working Conditions Based on Adaptive Regularized Transfer Learning [in Chinese], *Chinese Journal of Scientific Instrument*, **41**(08): 95-103, 2021, https://doi.org/10.19650/j.cnki.cjsi.J2107721
10 Zhou H., Wang Z., Tao Q., Rolling Bearing Fault Diagnosis Based on Joint Structure-Preserving Transfer of Multi-Source Domains [in Chinese], *Journal of Vibration and Shock*, **44**(14): 302-310, 2025.
11 Fu J., Zhang G., Zhang S., Bearing fault diagnosis of cross-working grinding mill based on transfer learning [in Chinese], *Journal of Fujian University of Technology*, **22**(04):393-400, 2024.
12 Pan X., Ge K., Dong F., Intelligent Fault Diagnosis of hoist bearing based on Feature Transfer Learning [in Chinese], *Industrial and Mine Automation*, **48**(09): 1-7+32, 2022, https://doi.org/10.13272/j.issn.1671-251x.17980
13 Wan A., Yang J., Wang J., Chen T., Liao X., Huang J., Du X., Aero engine Gear Fault Diagnosis Based on Deep Learning. Vibration [in Chinese], *Test and Diagnosis*, **42**(06): 1062-1067+1239, 2022, https://doi.org/10.16450/j.cnki.issn.1004-6801.2022.06.002
14 Hu W., Zhang Y., Centrifugal Pump Rolling Bearing Fault Diagnosis based on VMD and Random Forest [in Chinese], *Mechanical and Electrical Engineering Technology*, **51**(03):78-82, 2022.
15 Wang J., Zhou C., Zhang Z., Mend N., Du G., Jin X., Zhang C., A fault diagnosis method using decomposition denoising improved multiscale weighted permutation entropy and one-versus-one least squares twin SVM, *Measurement*, (255): 118012, 2025, https://doi.org/10.1016/j.measurement.2025.118012
16 Liu J., Maimaitireyimu A., Xiang Z., Xie L., Fan Bearing Fault Diagnosis Based on Improved Grey Wolf Optimization Algorithm and SVM, *Journal of Mechanical Transmission*, **47**(09):160-169, 2023, https://doi.org/10.16578/j.issn.1004.2539.2023.09.022
17 Bu C., Liu Y., Zhang W., Fault Diagnosis Model of Shearer Bearing Based on PSO-SVM [in Chinese], *Industrial and Mine Automation*, **51**(S1): 44-46, 2025.
18 Guo Y., Liu Y., Zhang Z., Wang Y., Xue P., Du C., Li W., Research on fault detection and diagnosis of carbon dioxide heat pump systems in buildings based on transfer learning, *Journal of Building Engineering*, (85):108774, 2024, https://doi.org/10.1016/j.jobe.2024.108774
19 Wang B., Liu Y., Liao Y., Sensitivity analysis of time-domain characteristic indicators of rolling bearing fault signals, *Bearing*, (10): 45-48, 2015, https://doi.org/10.19533/j.issn1000-3762.2015.10.014
20 Tang J., Wang E., Zhu J., Tan W., Research on Fault Diagnosis of Automotive Water Pump Bearings Based on Frequency Domain Features and Support Vector Machines [in Chinese], *Machine Tool and Hydraulic*, **46**(13): 163-167+155, 2018, https://doi.org/10.3969/j.issn.1001－3881.2018.13.040
21 Li D., Feng J., Fault Diagnosis of Rolling Bearings in Medical Devices Based on Vibration Signal Analysis [in Chinese], *Value Engineering*, **44**(24): 21-23, 2025, https://doi.org/10.3969/j.issn.1006-4311.2025.24.007
22 Wang T., Liu T., Liu Y., Wang Z., Bearing Fault Transfer Diagnosis with Balanced Distribution Adaptation Under Variable Working Conditions [in Chinese], *Mechanical Science and Technology*, **42**(08): 1316-1323, 2023, https://doi.org/10.13433/j.cnki.1003-8728.20220058
23 Lei Y., Yang B., Jiang X., Jia F., Li N., Nandi A.K., Applications of machine learning to machine fault diagnosis: A review and roadmap, *Mechanical Systems and Signal Processing*, **138**: 106587, 2020, https://doi.org/10.1016/j.ymssp.2019.106587
24 Weiss K., Khoshgoftaar T.M., Wang D., A survey of transfer learning, *Journal of Big Data*, **3**(1): 1-40, 2016. https://doi.org/10.1186/s40537-016-0043-6
25 Wang J., Chen Y., Hao S., Feng W., Shen Z., Balanced Distribution Adaptation for Transfer Learning, *IEEE International Conference on Data Mining (ICDM)*, 1129-1134, 2017, https://doi.org/10.1109/ICDM.2017.150
26 Cortes C., Vapnik V., Support-vector networks, *Machine Learning*, **20**: 273 – 297, 1995, https://doi.org/10.1007/BF00994018
27 Schölkopf B., Smola A.J., Learning with Kernels, *MIT Press*, 2002. https://doi.org/10.7551/mitpress/4175.001.0001