ONLINE FIRST June 9, 2026

Engineering Transactions, **74**, https://doi.org/10.24423/engtrans.2026.3649
Institute of Fundamental Technological Research Polish Academy of Sciences (IPPT PAN)
Université de Lorraine • Poznan University of Technology

Research Paper

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, China

*Corresponding Author: chendong859052@163.com

To address the issue of low diagnostic accuracy caused by distribution differences between the source and target domains in rolling bearing fault diagnosis, this study proposes a method combining balanced distribution adaptation (BDA) and support vector machines (SVMs). The approach utilizes BDA to simultaneously minimize discrepancies in both the marginal 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 the feature distributions, and SVMs are employed for fault classification. Experimental results demonstrate that the BDA-SVM method achieves over 94 % diagnostic accuracy, showcasing strong performance and robustness in bearing fault diagnosis. Compared with traditional SVMs and other methods without transfer learning, the proposed approach shows a significant improvement in diagnostic accuracy under cross-domain conditions.

Keywords: rolling bearings, fault diagnosis, balanced distribution adaptation, transfer learning, support vector machines (SVMs).

1. Introduction

Bearings in industrial equipment generally play a critical role in support and power 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 escalation [5].

Traditional fault diagnosis relies on manual experience analysis, which is inefficient, highly subjective, and requires precise mechanistic models as support, thus presenting numerous challenges in practical applications [6]. With technological advancements and the accumulation of experience, data-driven intelligent diagnostic methods, by 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 data collected under normal operating conditions are relatively abundant in most industrial scenarios, fault data are often scarce due to various challenges such as the difficulty of fault 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 Case Western Reserve University (CWRU) bearing dataset, including balanced distribution adaptation (BDA)+K-nearest neighbor (KNN) and BDA+support vector machines (SVMs), and reported that BDA+KNN achieved the highest accuracy, while BDA+SVMs ranked second. Similarly, Lei *et al.* [8] provided a comprehensive review of deep-learning-based fault diagnosis methods and highlighted their effectiveness in feature extraction under varying working conditions. Although these deep learning-based methods have shown promising performance, they often require a larger number of parameters and substantial 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 objective is to reduce the distribution differences between the source and target domains to achieve knowledge transfer, enabling the model to effectively reuse source-domain knowledge in the target domain [7]. Existing studies mostly focus on marginal distribution features between the source and target domains while neglecting conditional distribution differences [1, 8]. BDA achieves effective cross-domain feature transfer by jointly optimizing differences between marginal and conditional distributions and introducing a tuning parameter to balance the weights of marginal and conditional distributions. This approach preserves category-discriminative information while minimizing domain-specific bias, thereby providing an effective

method for fault diagnosis under different operating conditions and across different devices [8–10].

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

Based on the above considerations, this study proposes a rolling bearing fault diagnosis method combining BDA and SVM: the BDA algorithm is used to jointly align the marginal and conditional distributions of the source and target domains, thereby addressing the limitation of existing studies that only consider marginal distributions to achieve feature-level adaptation. Meanwhile, 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-domain and small-sample conditions.

2. Signal time- and 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 the mean value, root mean square (RMS), spectral energy, etc.) is required to convert the signals into quantitative indicators that reflect the key features of these signals. These indicators are first aligned between the source domain and target domain using BDA, and are 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- and frequency-domain features as the inputs to BDA [22] and SVM.

2.1. 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 value, RMS, and variance are commonly used in signal analysis and also have practical significance in engineering applications.

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

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{2.1}$$

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

The 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

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}. \tag{2.2}$$

Skewness (S) is a statistical measure describing the symmetry of a signal and is used to analyze signal asymmetry. Its expression is

$$S = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \mu}{\sigma}\right)^3, \tag{2.3}$$

where 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-related faults. Its expression is

$$K = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \mu}{\sigma}\right)^4. \tag{2.4}$$

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

$$V = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2. \tag{2.5}$$

The crest factor (C) is defined as the ratio of the maximum value of the signal to its RMS value. It is typically related to the impacts experienced by the system and can be used to assess the intensity of transient impacts in the signal. Its expression is

$$C = \frac{\max(|x_i|)}{X_{\mathrm{RMS}}}. \tag{2.6}$$

2.2. 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 types of faults in rolling bearings have distinct characteristic frequencies and are closely related to the inner and outer rings, rolling element dimensions, contact conditions, and relative rotational frequency [19]. In this paper, we extract two common frequency-domain features: spectral center frequency and spectral energy.

The spectral center frequency (CF) is used to describe the central position of a frequency-domain signal, i.e., the average value of the signal frequency distribution, and can indicate the dominant frequency where the signal energy is concentrated. Its expression is

$$\mathrm{CF} = \frac{\sum_{i=1}^{N} f_i P(f_i)}{\sum_{i=1}^{N} P(f_i)}, \tag{2.7}$$

where f_i is the frequency component of the signal, and $P(f_i)$ is the corresponding power spectral density.

Faults cause a significant increase in energy at the characteristic frequency locations [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

$$\mathrm{SE} = \sum_{i=1}^{N} |X(f_i)|^2, \tag{2.8}$$

where $X(f_i)$ is the amplitude of the Fourier transform spectrum of the signal.

3. BDA-SVM algorithm principle

3.1. 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 and target domains [7–9]. The objective 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 from the source domain (such as model parameters, feature distributions, or latent-space representations) to improve the performance of tasks in the target domain, enabling models trained on the source domain to be effectively applied to the target domain after fine-tuning.

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

$$D_s = \left\{ (x_s^{(i)}, y_s^{(i)}) \right\}_{i=1}^{n_s}, \tag{3.1}$$

$$D_t = \left\{ x_t^{(j)} \right\}_{j=1}^{n_t}, \tag{3.2}$$

where 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 numbers of samples in the source and target domains, 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. Therefore, we employ pseudo-labeling techniques to estimate the conditional distribution, where initial predictions from the source domain model are used to approximate the true target-domain labels.

The feature distributions of the source and target domains $P_S(X)$ and $P_T(X)$, as well as the joint distributions $P_S(XY)$ and $P_T(XY)$, are typically different. Therefore, the objective of transfer learning is to train a model $f_S(x)$ in the source domain that can also 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.

3.2. Introduction to BDA algorithms

BDA is a domain adaptation technique that simultaneously minimizes differences between marginal and conditional distributions, thereby aligning the source and target domains in terms of feature space 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 over the entire feature space, whereas the CDD considers the distribution differences between samples belonging to the same category in the source and target domains, thereby ensuring that

classification-related information remains consistent during the transfer process. The expressions are as follows:

$$\text{MDD} = \left\| \mathrm{E}_{x_i^S \sim P_S(X)} \left[\Phi(x_i^S) \right] - \mathrm{E}_{x_i^T \sim P_T(X)} \left[\Phi(x_i^T) \right] \right\|^2, \tag{3.3}$$

$$\text{CDD} = \sum_{c \in C} \left\| \mathrm{E}_{x_i^S \sim P_S(X|Y=c)} \left[\Phi(x_i^S) \right] - \mathrm{E}_{x_i^T \sim P_T(X|Y=c)} \left[\Phi(x_i^T) \right] \right\|^2. \tag{3.4}$$

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 $P(y^T|x^T)$. It is important to note that although 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 feature 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 feature alignment improves, leading to more accurate conditional distribution matching. This approach allows BDA to effectively handle 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:

$$L_{\text{BDA}} = \text{MDD} + \lambda \text{CDD}, \tag{3.5}$$

where λ is a hyperparameter used to adjust the influence of marginal and conditional distributions during transfer learning. By adjusting λ, the model's adaptability to distribution differences between the source and target domains can be controlled. The core idea of BDA is to simultaneously coordinate the optimization of MDD and CDD by introducing the balancing parameter λ.

3.3. 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], while the kernel trick for nonlinear mapping follows the formulation proposed by Schölkopf and Smola [27]. The core idea of SVM is to maximize the separation margin between samples from 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-versus-rest or one-versus-one. In this study, we employ the one-versus-rest approach for the four-class bearing fault diagnosis task. Assume that the data points in the dataset are $(x_i y_i)$, where $x_i \in R^n$ denotes the input feature vectors, andy_i represents the corresponding class label. The SVM classification hyperplane can be expressed as follows:

$$d = \frac{\left|w^T x_i + b\right|}{\|w\|}, \tag{3.6}$$

where 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

$$\frac{y_i(w^T x + b)}{\|w\|}. \tag{3.7}$$

To maximize this margin while ensuring correct classification, the concept of a soft margin is introduced to allow certain sample points to lie on or inside the decision boundary, thereby determining the final position of the classification hyperplane. To find the optimal classification hyperplane, SVM solves the following optimization problem:

$$\min_{w,b} \frac{1}{2} \|w\|^2, \tag{3.8}$$

under the constraints:

$$y_i(w^T x_i + b) \geq 1, \qquad i = 1, 2, ..., n, \tag{3.9}$$

where $\frac{1}{2} \|w\|^2$ is the regularization term, which prevents overfitting and ensures that the classifier has good generalization ability.

However, in practical applications, many datasets cannot be well separated by a single linear hyperplane. To address this issue, SVM introduces kernel functions that map the input data into a higher-dimensional feature space, thereby achieving linear separability in high-dimensional spaces. Since the radial basis function (RBF) kernel effectively handles nonlinear classification problems and exhibits strong classification capabilities in high-dimensional spaces, it is selected as the built-in kernel function for the SVM model in this paper, and its expression is as follows:

$$K(x_i x_j) = \exp(-\gamma \|x_i - x_j\|^2). \tag{3.10}$$

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

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i, \tag{3.11}$$

where C is the penalty parameter that controls the trade-off between margin maximization 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

$$f(z) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(z_i, z) + b\right), \tag{3.12}$$

where α_i is the Lagrange multiplier obtained by solving the optimization problem.

3.4. BDA-SVM rolling bearing fault diagnosis strategy

By combining the advantages of BDA and SVM, this study proposes a BDA-SVM-based rolling bearing fault diagnosis strategy to address fault diagnosis problems when there is significant feature distribution discrepancy between the source and target domains. The workflow of the proposed model is shown in Fig. 1. The white boxes represent the inputs and outputs, while the black boxes represent the intermediate processes.

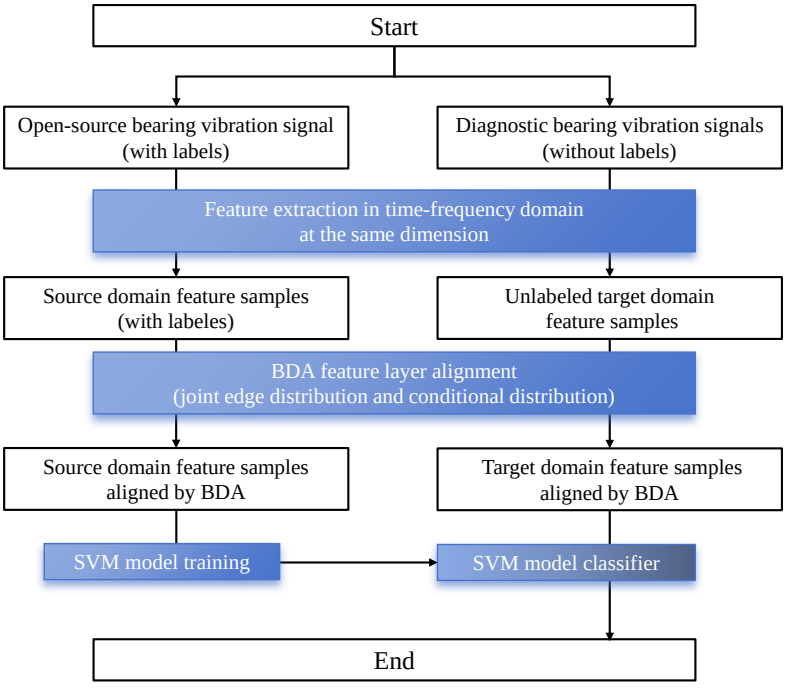

Fig. 1. BDA-SVM fault diagnosis flowchart.

As shown in Fig. 1, first vibration signals from open-source datasets and rolling bearing fault vibration signals are acquired. Time-domain and frequency-domain feature extraction are then performed on these two types of signals. The extracted time-domain features used in this study typically include the mean value, 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 and target domains. Finally, the transformed feature data are input into the SVM classifier to train the fault classification and diagnosis model. The SVM model is then trained and performs classification on the input target-domain features to identify different fault types and ultimately complete the fault diagnosis task.

4. Experiments and analysis

4.1. 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 the experimental data. The vibration signals were sampled at 12 kHz with a duration of 1 s per sample, resulting in 12 000 data points for each signal segment. Sequential sampling was used without overlap between adjacent segments. To validate the effectiveness of the proposed method for rolling bearing fault diagnosis, the CWRU bearing dataset was used. The CWRU dataset includes four main fault modes: normal condition, inner ring fault, outer ring fault, and rolling element fault. The fault category codes are shown in Table 1.

Table 1. Bearing dataset fault labels.

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

Inner ring faults manifest as physical damage on the inner race of the bearing, while outer ring faults are associated with damage to the outer race. Rolling element faults indicate abnormal damage to the balls or rollers. Each fault category includes different damage levels (0.007 inch, 0.014 inch, and 0.021 inch).

The data are collected by recording vibration signals during bearing operation using an accelerometer.

Additionally, the experiment utilizes rolling bearing fault data collected from actual industrial equipment. These data serve as the target domain to validate the model's transferability in real-world industrial applications. The experimental setup, shown in Fig. 2, employed SKF 6205 deep-groove ball bearings. Faults were artificially introduced using electrical-discharge machining to create single-point defects on the inner race, outer race, and rolling elements, with fault diameters of 0.007 inch, 0.014 inch, and 0.021 inch, 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 Fig. 2.

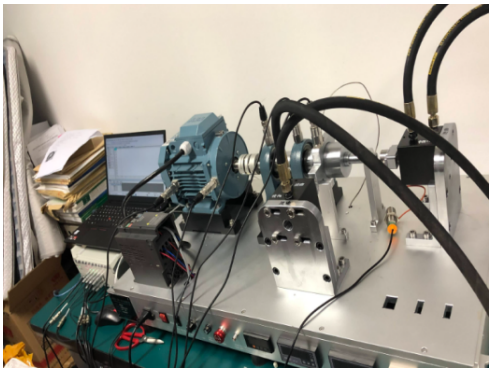

Fig. 2. Bearing signal acquisition device.

4.2. Time- and frequency-domain responses of normal and faulty bearings

The signals acquired by the device are shown in Fig. 3. 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 Fig. 3. By comparing different fault conditions, it can be observed that any fault types cause 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 become significantly more pronounced, indicating the validity of using time- and frequency-domain features for fault characterization.

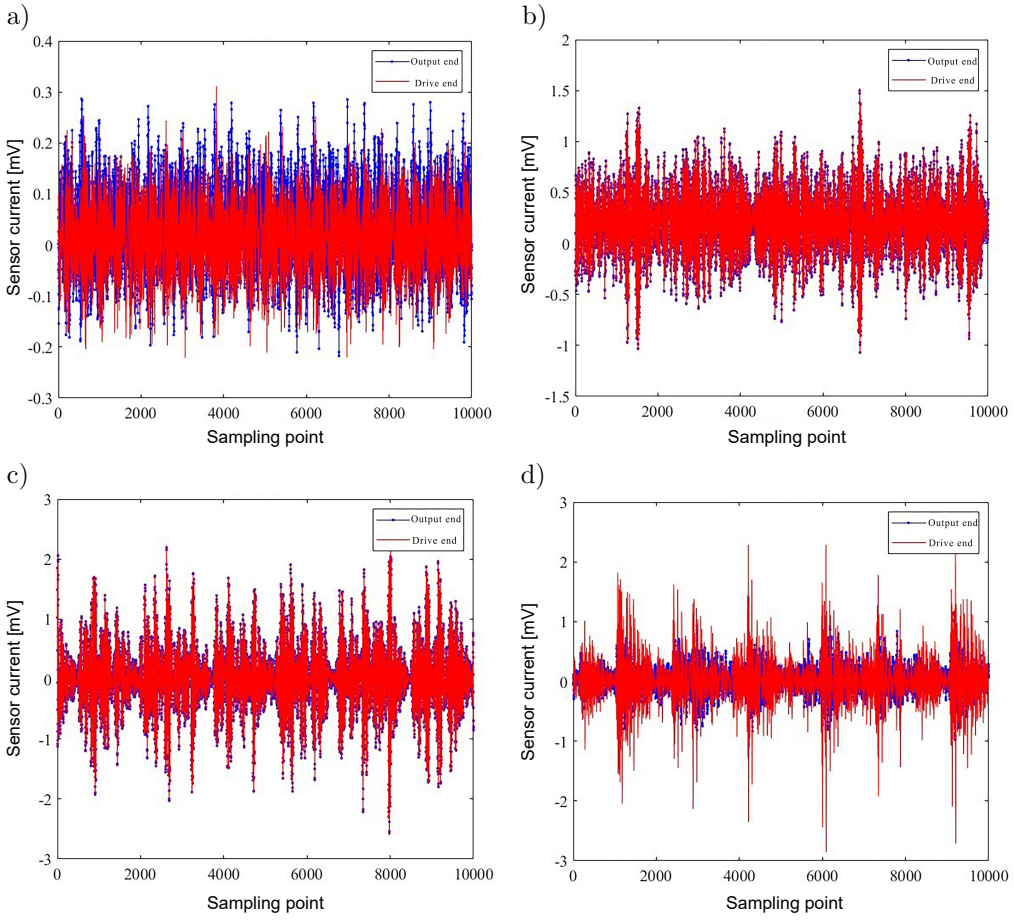

FIG. 3. Time-domain signals collected in the experiment: a) normal condition, b) rolling element fault, c) inner ring fault, d) outer ring fault.

To further investigate the frequency components of the signal, we normalized the frequency-domain features, and the resulting frequency-domain signal is shown in Fig. 4.

As shown in Fig. 4a, 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, while wide-band frequency characteristics are not prominent. Under rolling element fault conditions, the amplitude near the characteristic frequency (0.12) of the rolling elements is significantly higher than that of the fundamental frequency (0.17). Under inner ring fault 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 condi-

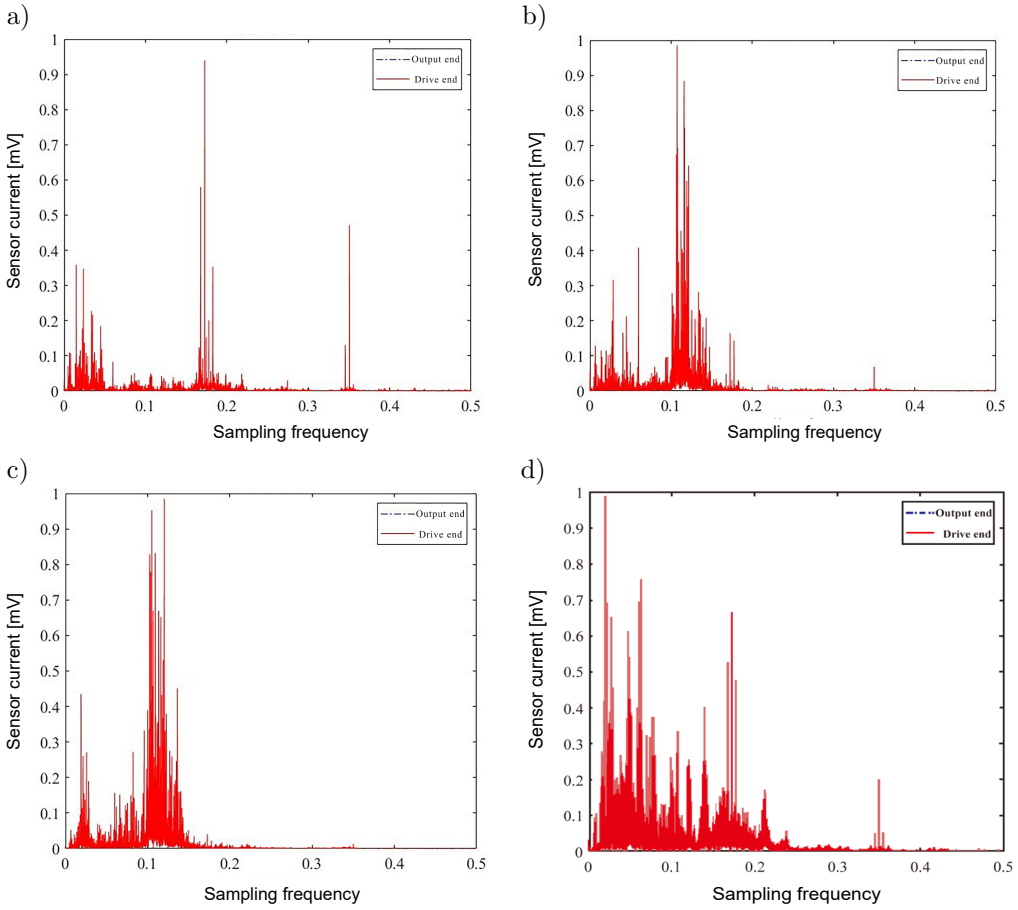

FIG. 4. Frequency-domain features of signals acquired in the experiment: a) normal condition, b) rolling element fault, c) inner ring fault, d) outer ring fault.

tions, the wide-band frequency characteristics are more pronounced, and the frequency components of 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 Fig. 3d, the fault signals are easier to identify in the time-domain signals compared with the frequency-domain signals in Fig. 4d. Therefore, using combined time- and frequency-domain characteristics as diagnostic indicators is more reasonable.

4.3. Validation of the BDA method

In this experiment, the source domain consists of the CWRU dataset under specific operating conditions, while the target domain comprises collected

under 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 employ a stratified random sampling strategy to divide the datasets. To achieve effective transfer learning, 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 minimization of feature distribution differences.

In actual bearing fault diagnosis, the source 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 and target domains, 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 5 shows the visualization results of BDA and TCA before and after feature transfer.

As can be clearly seen from the classification results in Fig. 5a, before feature migration, due to the significant differences in data distribution between the source domain and target domains, data points from different categories overlap in the feature space, resulting in blurred classification boundaries and making it difficult to perform effective fault classification. After applying the TCA method for feature transfer (Fig. 5b), the distribution of sample data from

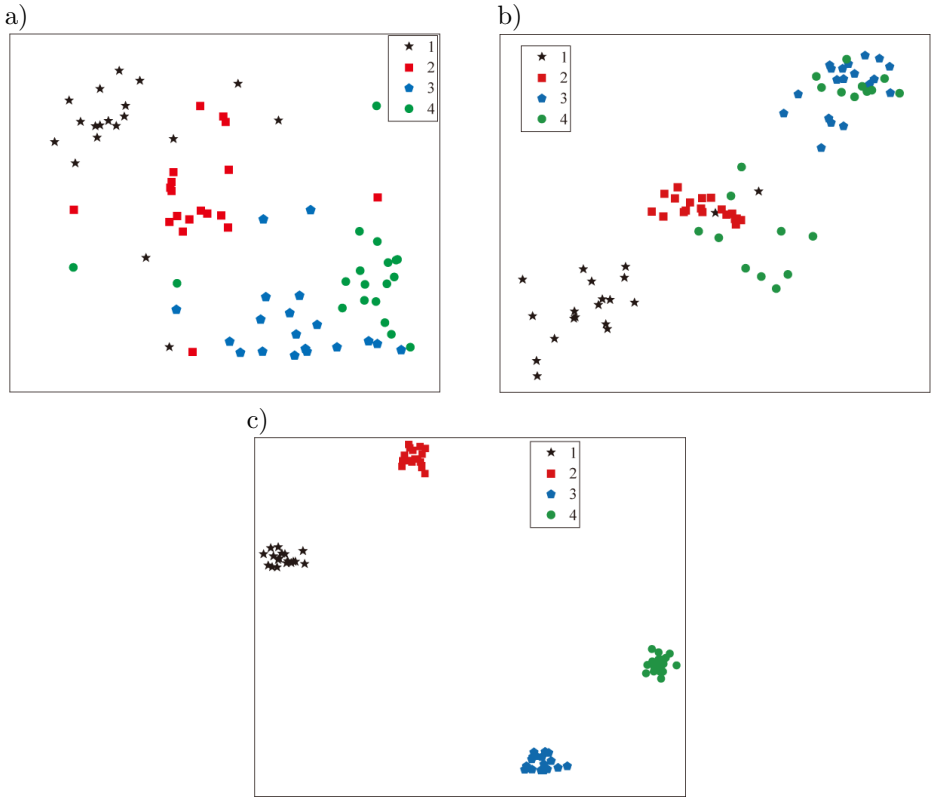

FIG. 5. Visualization results before and after feature transfer: a) before feature transfer, b) after TCA-based feature transfer, c) after BDA-based feature transfer.

different fault categories in the feature space improves, with some categories beginning to cluster together and distribute more densely. This is because TCA can reduce the distribution differences between the source and target domains. 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 account for imbalances in feature distributions between the source domain and target domains, resulting in insufficient separation of some data categories and thereby affecting classification accuracy.

After feature transfer using the BDA method (Fig. 5c), the data distribution of the source domain and target domain are significantly optimized. Data points of different fault categories become more compact and distinct in the feature space, and the distribution differences between the source and target domains are greatly reduced. BDA achieves this by adaptively adjusting the feature distributions between the source and target domains, making samples from the same category to be more consistently distributed across domains while enhancing the separation between categories. This significantly improves classification

accuracy and stability. These results indicate 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 distributions of the source and target domain data. The original feature space consists of the nine time-frequency domain features extracted in Sec. 1 (including mean, RMS, skewness, kurtosis, variance, crest factor, spectral center frequency, and spectral energy). To visualize these high-dimensional features in a 2D plane, we employ 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.

4.4. Performance validation of the BDA-SVM model

To validate the performance of the proposed BDA-SVM model, this paper designed a series of comparative experiments to compare the diagnostic effectiveness of several traditional classification algorithms, including decision trees (DT), 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 Fig. 6.

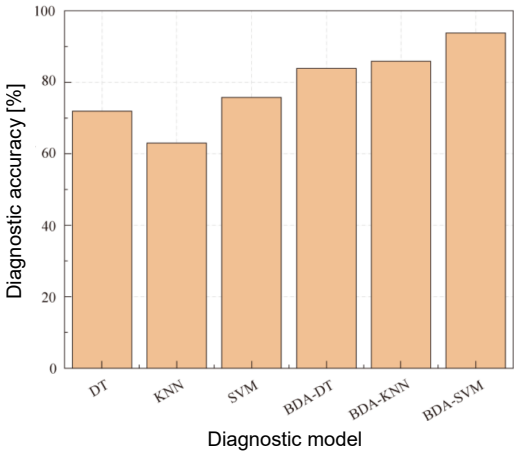

Fig. 6. Diagnostic results of different models.

As can be seen from the comparison results in Fig. 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.

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 improvement is attributed to the substantial reduction in distribution differences between the source and target domains after BDA processing, enabling the SVM to better adapt to the target-domain data. Compared with the other 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. These results indicate 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 characterized by 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, as shown in Fig. 6.

5. Conclusion

This paper addressed the distribution differences between the source and target domains in cross-domain fault diagnosis and proposed a rolling bearing fault diagnosis strategy based on BDA-SVM. The BDA algorithm effectively

reduced the distribution differences between the source and target domains at the feature level, thereby enhancing the model's generalization capability 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 %, and 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 still has certain limitations. In particular, 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.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that there are no known competing financial interests or personal relationships that could have influenced the work described in this paper.

Authors' contributions

Dong Chen conceptualized the study and wrote the original draft. Mingshan Zhang conducted experiments,performed the analysis and contributed to data interpretation. Yaguang Gao performed the analysis and wrote the original draft. All authors reviewed and approved the final manuscript.

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., Xu X., Han X., 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**(Part A): 113727, 2025, https://doi.org/10.1016/j.asoc.2025.113727.
7. Ma J., Li C., Zhang G., Rolling bearing fault diagnosis based on deep learning and autoencoder information fusion, *Symmetry*, **14**(1): 13, 2021, https://doi.org/10.3390/sym14010013.
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*, **42**(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., Meng 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., Maimaitireyim 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 [in Chinese], *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, [in:] *2017 IEEE International Conference on Data Mining (ICDM)*, pp. 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: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2002, https://doi.org/10.7551/mitpress/4175.001.0001.

Received August 31, 2025; accepted December 2, 2025;
available online April 7, 2026; version of record June 9, 2026;
published issue XXXX.