Cross-domain intelligent fault diagnosis of rolling bearing based on distance metric transfer learning

Hongdi Zhou, Tao Huang, Xixing Li and Fei Zhong

Abstract
Rolling bearings are present ubiquitously in mechanical equipment, timely fault diagnosis has great significance in guaranteeing the safety of mechanical operation. In real world industrial applications, the distribution of training dataset (source domain) and testing dataset (target domain) is often different and varies with operating environment, which may lead to performance degradation. In this study, a cross-domain fault diagnosis of rolling bearing method based on distance metric transfer learning (DMTL) and wavelet packet decomposition (WPD) is proposed. The Mahalanobis distance is adopted for learning the intrinsic similarity or dissimilarity between instances and learned by simultaneously maximizing the intra-class distances and minimizing the inter-class distances for target domain. The features of source domain and target domain are first extracted from original vibration signals by WPD which is a powerful tool in dealing with non-stationary signals and can provide meticulous analysis. Then, the DMTL model is adopted to eliminate the error propagation across different components, which can weaken the weight of low-quality instances and enhance the weight of high-quality samples. Finally, the k-nearest neighbor (KNN) classifier is applied to accomplish the cross-domain intelligent fault-type classification. The superiority and effectiveness of the proposed fault diagnosis model is validated by two diagnosis cases. The experimental results demonstrated that the proposed method performs better than other compared methods in recognizing various fault types and has the capability in handling the complex cross-domain adaptation scenarios.

Keywords
Bearing fault diagnosis, distance metric transfer learning, wavelet packet decomposition, Mahalanobis distance

Introduction
Rolling bearings play an indispensable role in equipment, which are prone to breakdown since they often operate with awful conditions, such as high temperature, heavy loads, high rotating speed, and etc., and almost 45%–55% of rotating machinery failures are rolling bearing faults.1–3 Unexpected failures may boost the cost of operation, maintenance and even lead to catastrophic casualties.4 To ensure the safety and reliability of the rotating machinery, accurate and efficient diagnosis of incipient faults is extremely important.

Conventionally, the fault diagnostic techniques collect and process various signals with the goals of resuming from malfunctions or faults and precluding from future failures as early as possible.2,5 Data-driven fault recognition approaches related to artificial intelligence techniques or machine learning techniques, such as neural networks, support vector machines, and decision tree methods, are extensively used for bearing fault diagnosis. However, these methods are more dependent on the data characteristics and distribution, which limits their adaptability to the cross-domain scenarios. In order to address this limitation, transfer learning methods are introduced to extend the performance of existing methods to other domains.

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
as support vector machine (SVM), k-nearest neighbor (KNN), and artificial neural network (ANN), etc., have been extensively studied to improve existing techniques with the goal of more accurately and effectively dealing with various complex problems, such as varying load effect and noise contamination.6–8 Additionally, deep learning methods have been widely used for condition recognition over the past decades.9–11 These intelligent recognition techniques have achieved great success in distinguishing operating conditions of various machines under complex working environment.

Despite huge success, most of the intelligent recognition methods work well under two general hypotheses: one is abundant labeled training samples are available; the other is the training and testing dataset are drawn from the identically probability distribution. However, in real-world scenarios, the performances of these methods may dramatically decline because of the variations between the distributions of training data (source domain) and testing data (target domain).12 The distribution of collecting dataset varies with the operating environment, such as the installation conditions of experimental platform, motor loads, humidity, temperature, and etc., which is known as cross-domain learning problem.13 The variations (domain shifts) could cause a great discrepancy between the features extracted from the signals obtained from experimental settings and the signals collected from actual operating situation.

Recently, transfer learning methods have received extensive studies, which can adapt a machine learning model trained by dataset of source domain to a different but related target domain.14–16 The learning strategies of most published transfer learning methods can be roughly divided into three categories: feature-based methods, instance based methods and metric-based methods.17 The feature-based learning aims to discover a feature subspace in which the recognition model trained in source domain is qualified for target domain.18,19 Instance-based transfer learning aims at reweighting the source samples according to the shared information provided by target data, which the reweighted instances can be further analyzed.20–22 Metric-based methods, distance metric learning (DML) algorithms, aims to learn an optimal distance metric for measuring sample pairs similarity or dissimilarity by exploiting meaningful correlations between instances in source and target domain,17 which can effectively reduce the distribution divergence between domains and extract the weakly discriminative information.23 Cao et al. proposed consistent distance metric learning to estimate the instance weights under covariate shift situations, which the Euclidean distance metric was utilized to determine sample pairs correlation.24 Huang and Zhou presented an unsupervised metric transfer learning method (UMTL) to learn domain invariant features with more discriminative information via Maximum Mean Discrepancy.25 Most existing transfer learning algorithms use the Euclidean distance to estimate the dissimilarity or similarity between the samples in source or target domain.

However, the Euclidean distance is adopted to measure the dissimilarity and similarity between samples in most existing transfer learning techniques, which may decline the transfer learning performance since Euclidean distance would not maximize the inter-class distances while minimizing intra-class distances.26 Ahmadvand and Tahmoresnezhad presented Metric Transfer Learning via Geometric Knowledge Embedding, Mahalanobis distance metric and the graph optimization were employed to reweight the instance weights of source samples for distribution matching.17 Xu et al. put forward a metric transfer learning framework to encode metric learning in transfer learning.27 Therefore, Mahalanobis-distance-based transfer learning can effectively minimize the distance between source and target domains.

Inspired by the strategy of transferring knowledge from source domain to target domain, a cross-domain fault diagnosis of rolling bearing method based on distance metric transfer learning (DMTL) and wavelet packet decomposition (WPD) is proposed in this study. Based on the metric learning, the Mahalanobis distance instead of Euclidean distance is adopted in the objective function of cross-domain adaption for learning the intrinsic similarity or dissimilarity between instances. Then, the intra-class distances of target domain are maximized and the inter-class distances of target domain are minimized to improve recognition accuracy by using the intrinsic information among samples from different domains with labels. Experimental investigations are carried out to demonstrate the feasibility and effectiveness of the proposed method for rolling bearing fault diagnosis.

The rest of the article is organized as follows. In section II, the principle of DMTL is introduced. Then, the fault diagnosis model based on the DMTL and WPD algorithm is presented in Section III. After that, the practical cases are studied to validate the superior performance of the proposed model in Section IV. Finally, some concluding remarks are summarized in Section V.

**Principle of distance metric transfer learning**

Suppose the domain \( D = \{X, P(X)\} \) is composed of a \( d \)-dimensional feature space \( X \) and a marginal probability distribution \( P(X) \), where \( X = \{x_i\}_{i=1}^{n} \in \mathbb{R}^{d \times n} \) is a dataset consists of each \( x_i \in \mathbb{R}^{d \times 1} \) sample from this domain. Its corresponding task \( T = \{Y, f(X)\} \) is
composed of a label space $Y$ and a prediction function $f(X)$ where $Y = \{y_i\}_{i=1}^{N_Y}$ is the label vector of feature dataset $X$ with $y_i$ is the label of $x_i$, and $f(X) = Q(Y|X)$ is the conditional probability distribution. Let $D_s = \{X_s, P_s(X_s)\}$ and $D_t = \{X_t, P_t(X_t)\}$ denote the source domain and target domain respectively, their corresponding feature dataset are $X_s = \{x_{s:i}\}_{i=1}^{N_X}$ and $X_t = \{x_{t:i}\}_{i=1}^{N_X}$, and the samples of source domain are labeled and only a few samples of target domain are labeled. Let $T_s = \{Y_s, Q(Y_s|X_s)\}$ and $T_t = \{Y_t, Q(Y_t|X_t)\}$ denote source task and target task, respectively. This study is focus on the homogeneous metric transfer learning, which implies that the feature space and label space of source domain and target domain are same, while the marginal probability distribution and the conditional probability distribution are different, that is, $X_s = X_t$, $Y_s = Y_t$, $P_s(X_s) \neq P_t(X_t)$ and $Q(Y_s|X_s) \neq Q(Y_t|X_t)$. The experimental dataset contains $C$ different operating statuses and each status has $n^c$ samples in the subdomain. Without loss of generality, let $Y_s = Y_t = \{1, \ldots, C\}$ and let $y_i = 1$ represents the normal status of rotating machinery in fault diagnosis.

Since there is a discrepancy between the distribution of source domain and target domain, the labeled samples of source domain cannot be directly applied to learn a distance metric for target domain. To address this issue, the labeled samples of source domain are reweighted meanwhile preserving the distance relation among data in source domain and target domain, which can provide discriminative information for target domain. In this study, a reweighting instance strategy called distance metric transfer learning (DMTL) is investigated, and the objective function of DMTL method consists three parts as follows:

$$\min_{A, w} R = R_A + \lambda R_w + \beta R_j$$

where the first term $R_A$ is the primary objective of distance metric learning as same with distance metric learning, which controls the generalization error of the distance metric. The second term $R_w$ is the regularization term on instance weights of source-domain labeled samples. The third term $R_j$ is the loss function of prediction model of the learned distance metric on target domain labeled samples along with the reweighted source-domain labeled samples. $\lambda > 0$ and $\beta > 0$ are the trade-off parameters to balance impact of those three terms in equation (1).

**Regularization term of instance weights $R_w$**

To avoid the potential issues in studying the instance weights $w$ and Mahalanobis distance metric $A$, the regularization term of instance weights $R_w$ is applied, which can effectively estimate the instance weights. The $R_w$ term is defined as follows:

$$R_w = ||w(x) - w_0(x)||^2$$

where $w_0(x) = \frac{P_s(x)}{P_s(x) + P_t(x)}$ are the estimated density ratios or weights of instances $x$ of source domain with Euclidean metric, $w(x) = \frac{P_t^A(x)}{P_s^A(x)}$ is the ideal instance weights with Mahalanobis distance metric, $A$ and $A'$ are the ideal distance metric of source domain and target domain, respectively; $P_t^A(x)$ and $P_s^A(x)$ are the density estimations of with distance metric $A$ and $A'$, respectively. Obviously, the higher value of $w_0(x)$, the higher value of $P_t(x)$ and the smaller value $P_s(x)$, which implies that $x$ is closer to distribution of target domain than that of source domain, and the instance weights of target domain are 1, that is, $w_0(x_T) = 1$ for $x_T \in D_T$.

The density ratios $w_0(x) = \frac{P_t(x)}{P_s(x) + P_t(x)}$ can be estimated by a linear combination of some basic functions, that is, $w_0(x) = \sum_{i=1}^{m} \alpha_i \phi_i(x)$, where $\phi_i$ represent a set of predefined basis functions and $\alpha_i$ is the corresponding positive parameters to be learned. The estimation performance of density ratios is determined by the
setting of $\varphi_i$. Thus, the Gaussian kernel function centered at $c_i$ is adopted to define the basic function $\psi_i(x) = \exp\left\{ -\frac{|x-c_i|^2}{\sigma^2} \right\}$, where $c_i$ can be calculated by a subset of instances of target domain. The weights of instances $x$ of source domain with Euclidean metric $w_0(x)$ can be obtained by minimizing KL-divergence between $P_T(x)$ and $w_0(x)P_S(x)$

$$\min_{w_0} KL(P_T(x) \parallel w_0(x)P_S(x))$$

$$= \int P_T(x) \log \frac{P_T(x)}{w_0(x)P_S(x)} \, dx$$

$$= \int P_T(x) \log \frac{P_T(x)}{P_S(x)} \, dx - \int P_T(x) \log w_0(x) \, dx$$

Here, equation (5) can be transformed by the optimization problem as follows:

$$\max_a \sum_{x \in D_T} \log \sum_{i=1}^n \alpha_i \psi_i(x)$$

$$s.t. \sum_{x \in D_T} \log \sum_{i=1}^n \alpha_i \psi_i(x) = N_S, \text{ and } \alpha > 0.$$ 

Since the logarithmic function is convex, the optimal solution of equation (6) can be calculated by gradient ascent approaches. As seen in equation (6), all the samples of source domain and target domain including labeled and unlabeled are adopted to deduce parameters $\alpha_i$. Then the initial weights of instances in source domain can be obtained through $w_0(x) = \frac{\sum_{i=1}^n \alpha_i \psi_i(x)}{\sum_{i=1}^n \alpha_i \psi_i(x)}$, which can be utilized to further deduce more precise weights $w(x)$ with regularization term $R_t$ in equation (1). Therefore, the distribution discrepancy between reweighted instances of source domain and instances of target domain can be minimized.

To improve the performance of knowledge transferred across domains, the Mahalanobis distance metric in terms of $A$ and $A'$ are introduced. The optimal solution of $w(x) = \frac{P_T(x)A}{P_S(x)A'}$ is similar to that of $w_0(x)$, which is resolved by minimizing KL divergence $\min_{w} KL\left(P_T(x) \parallel w(x)P_S(x)\right)$, and the $w_0(x)$ is applied as the initial weights to learn $w(x)$.27

**Loss function of prediction model $R_t$**

The loss function of prediction model with learned Mahalanobis distance metric $A$ of labeled samples in target domain along with the re-weighted labeled samples in source domain is introduced to improve the classification performance, which is defined by K nearest neighbor along with instance weights $w(x)$ under Mahalanobis distance metric $A$ as follows:

$$R_t = l_w(A, w) - l_b(A, w)$$

where $l_w(A, w)$ and $l_b(A, w)$ are within-class and between-class of the accumulated weighted differences measured by distance metric $A$, respectively, which are defined as follows:

$$l_w(A, w) = \sum_{y_i = y_j} w(x_i)w(x_j) \left\| A(x_i - x_j) \right\|^2$$

$$l_b(A, w) = \sum_{y_i \neq y_j} w(x_i)w(x_j) \left\| A(x_i - x_j) \right\|^2$$

Substituting equations (3–4), and equations (7–8) into equation (1), the objective function of DMTL method can be obtained:

$$\min_{A, \psi} = tr(A^TA) + \lambda \left\| A(x_i - x_j) \right\|^2$$

$$+ \beta \left( \sum_{y_i \neq y_j} w(x_i)w(x_j) \left\| A(x_i - x_j) \right\|^2 \delta_{ij} \right)$$

s.t. $\sum_{i=1}^{N_T} w(x_i) = N_S, \text{ and } w(x_i) \geq 0$.

where $\delta_{ij}$ is an indicator function, $\delta_{ij} = 1$ for $y_i = y_j$, and $\delta_{ij} = -1$ for $y_i \neq y_j$. As seen in equation (9), the in-class and out-of-class instance pairs, denoted by $C$, are utilized to estimate the loss function. The optimal solution of equation (9) can be converted into as follows:

$$\min_{A, \psi} = \min_{A, \psi} R_A + \lambda \left\| A(x_i - x_j) \right\|^2$$

$$+ \beta \left( \sum_{y_i \neq y_j} w(x_i)w(x_j) \left\| A(x_i - x_j) \right\|^2 \delta_{ij} \right)$$

where $\theta$ is a nonnegative penalty coefficient, $e_t = 1$ for $t \leq N_S$, and $e_t = 0$ for $N_S < t \leq N_S + N_T$.

Since the equation (10) is non-parametric, an alternating optimization algorithm is introduced to learn metric $A$ and instance weights $w(x)$ alternately and iteratively. Fixed the metric $A$ at the $t$-th iteration, then update the value of $w_t(x)$ by gradient descent:

$$w_{t+1}(x) = w_t(x) - \mu \frac{\partial R}{\partial w} \bigg|_{w_t} = w_t(x) - 2\lambda (w(x) - w_0(x))$$

$$+ \beta \zeta + \mu_1 \left( \max(0, -w(x)) \right)$$

where $\zeta = \sum_{i,j} w(x_i) \left\| A(x_i - x_j) \right\|^2 \delta_{ij}$ and $\chi = sign(\max(0, -w(x_j)))$. After updating the value of $w_{t+1}(x)$, the metric $A_{t+1}$ is regenerated while the $w_{t+1}(x)$ is fixed:

$$A_{t+1} = A_t - \mu_2 \frac{\partial R}{\partial A} \bigg|_{A_t} = A_t - \mu_2$$

$$\left( 2\beta \sum_{i,j} w(x_i)w(x_j)A_t(x_i - x_j)(x_i - x_j)^T \delta_{ij} + 2A_t \right)$$

$$\left( 2\beta \sum_{i,j} w(x_i)w(x_j)A_t(x_i - x_j)(x_i - x_j)^T \delta_{ij} + 2A_t \right)$$
The metric $A$ and instance weights $w(x)$ are alternately and iteratively updated until the variation of objective function is less than the default threshold during the iteration procedure or the number of iterations has reached the maximum value, the metric $A$ and instance weights $w(x)$ will be confirmed. The initial distance metric $d_0$ is learned from source domain, and the initial weights of instances $w_0(x)$ in source domain can be obtained through Euclidean distance.

**Process of fault diagnosis by distance metric transfer learning**

The sensitive features of source domain and target domain are first extracted from original vibration signals by using a unified feature extractor. The vibration signals of fault bearings are generally non-stationary, and WPD is a powerful tool in dealing with non-stationary signals which can provide more meticulous analysis.\(^{28}\) WPD can effectively decompose a signal into both high- and mid-frequency information along with the corresponding frequency regions, which is widely used for fault diagnosis.\(^{6,28,29}\) Therefore, the features related to WPD including the relative energy in a wavelet packet node (REWPN) and the entropy in a wavelet packet node (EWPN) is extracted, where the REWPN denotes the normalized energy of the wavelet packets node, and the EWPN indicates the uncertainty of normalized coefficients of the wavelet packets node.\(^6\) For a given sample $x(n)$, the $j$-th wavelet packet coefficients of the $i$-th wavelet packet node is defined as $C_i^j$, then the REWPN and EWPN can be obtained as follows:

\[
REWPN(i) = \frac{\sum_{j=1}^{K} (C_i^j)^2}{\sum_{m=1}^{N} \sum_{j=1}^{K} (C_i^m)^2} \quad (13)
\]

\[
EWPN(i) = -\sum_{j=1}^{K} p'_j \log_2(p'_j) \quad (14)
\]

where $p'_j = (C_i^j)^2 / \sum_{j=1}^{K} (C_i^j)^2$, $N$ is the total number of wavelet packet nodes, and $K$ is the total number of wavelet packet coefficients in each wavelet packet node.

After the construction of feature set, a transfer learning strategy of improving predictive performance is put forward. Since DMTL can effectively eliminate the error propagation across different components, a fault diagnosis model based on the DMTL is proposed in this study, which contains model training stage and diagnosis stage, and the flowchart of proposed fault diagnosis method is shown in Figure 1. The heterogeneous features of source domain and target domain are first constructed from original vibration signals, which are subsequently normalized by $l_2$-norm. Then the optimal instance weights are obtained by DMTL which can weaken the weight of low-quality instances and enhance the weight of high-quality samples of source domain in training stage, and the KNN classifier is trained using adjusted samples of new subspace. The optimal weights of test samples are obtained by trained DMTL model in diagnosis stage, the obtained features are fed in trained KNN model to recognize the running state of rotating machinery. The corresponding decisions or control measures can be put forward by the classification results.

![Figure 1. Overall framework of the proposed fault diagnosis method.](image)

**Experimental and results**

**Experiment design and datasets**

To validate the performance of proposed method, two cross-domain roll bearing fault diagnosis scenarios are conducted. Rolling bearings are vulnerability components for rotation machinery, the frequent faults of rolling bearings are inner race fault, outer race fault and ball fault.\(^5\) In engineering practice, the status of bearings is monitored by using vibration signals and temperature of system. Then a real-world run-to-failure bearing fault diagnosis are conducted to further demonstrate the performance of the proposed method. The detailed description is shown in Table 1.

(1) **Bearing fault diagnosis:** three domains, denoted as Datasets A, B, C, are utilized in this scenario, in which Datasets A and B provided by Case Western Reserve University (CWRU) bearing data center and Dataset C is acquired from the Society for Machinery Failure Prevention Technology (MFPT). For Datasets A and B, the tested bearing is SKF6205-2RS deep groove ball bearings, the sampling frequency of vibration signals is 12 kHz, each sample contains 2400 points. Four different operating statuses, including ball fault (BF), inner race fault (IR), outer race fault (OR), and
normal state (N), are considered, in which the measured signals contain three fault severities (0.18 and 0.36 mm) for BF, IR, and OR statues. Dataset A is constituted by the vibration signals acquired from drive end of motor at 0 hp (1797 rpm). Dataset B is consisted of vibration signals under those four operating conditions at 3 hp (1730 rpm), and each fault type contains two severity levels signals. For Dataset C, the tested bearings are deep groove ball bearings, the geometric parameters are listed in the following: pitch diameter ($P_d$) = 31.62 mm, ball diameter ($D_b$) = 5.97 mm, numbers of ball ($Z$) = 8, the contact angle ($\alpha$) = 0°. Dataset C is constituted by three operating statuses: N, OR, and IR, and the vibration signals are collected under different motor loads, range from 250 to 300 lbs, with 48,828 Hz sampling rate. The rotational speed is 1500 rpm, each status contains 100 samples, and each sample contains 10,000 points. About 15% of the samples of different conditions in target domain are randomly selected for cross-domain training, and the remaining samples in target domain are adopted for testing. Based on Datasets A, B, C, six cross-domain fault diagnosis of roll bearing tasks are conducted, which contains four four-categories recognition tasks (A→B, A→C, B→A and B→C) and two three-categories recognition tasks (C→A and C→B) by removing BF samples because the Dataset C only includes N, IR and OR conditions, where → denotes the adaptation is implemented from left-side datasets (source domain) to right-side datasets (target domain).

(2) **Real-World Run-to-Failure Stage Diagnosis:**

The knowledge transfer performance from public datasets to real-world fault severity diagnosis is verified in this testing scenario. The CWRU datasets and MFPT dataset are adopted as the source domain, while the dataset (short for Hubt dataset) collected on the rotor rolling bearing and gearbox integrated fault test bench of Hubei University of Technology is utilized as the target domain. The experimental system is displayed in Figure 2, the tested bearings are the same as CWRU datasets (NSK 6205 deep groove ball rolling bearings), and the geometric parameters are listed in the following: pitch diameter ($P_d$) = 39.04 mm, ball diameter ($D_b$) = 7.94 mm, numbers of ball ($Z$) = 9, the contact angle ($\alpha$) = 0°. The vibration signals are collected under 600 N with 10,240 Hz sampling rate, and the rotational speed is 1800 rpm. Normal condition, inner race fault and outer race fault are introduced, each status contains 100 samples with length of 4096 points. To validate the knowledge transfer performance, three types of target domains including slight fault stage, moderate fault and serious fault are considered. In this scenario, nine cross-domain bearing fault severity

![Figure 2. Bearing test rig.](image-url)

**Table 1.** Details of datasets.

| Datasets   | Condition         | Label | Fault category | Number |
|------------|-------------------|-------|----------------|--------|
| CWRU-A     | 29.95 Hz (1797 rpm) | 1     | N              | 100    |
|            |                    | 2     | OR             | 2×50 = 100 |
|            |                    | 3     | IR             | 2×50 = 100 |
|            |                    | 4     | BF             | 2×50 = 100 |
| CWRU-B     | 28.83 Hz (1730 rpm) | 1     | N              | 100    |
|            |                    | 2     | OR             | 2×50 = 100 |
|            |                    | 3     | IR             | 2×50 = 100 |
|            |                    | 4     | BF             | 2×50 = 100 |
| MFPT-C     | 25 Hz–270 lbs      | 1     | N              | 100    |
|            | 25 Hz–300 lbs      | 2     | OR             | 100    |
|            | 25 Hz–250/300 lbs  | 3     | IR             | 100    |
| Hubt       | 30 Hz (1800 rpm)   | 1     | N              | 100    |
|            |                    | 2     | OR             | 100    |
|            |                    | 3     | IR             | 100    |
diagnosis tasks (A—slight, B—slight, C—slight, A—moderate, B—moderate, C—moderate, A—serious, B—serious, C—serious) are conducted.

**Experimental results**

The sensitive features of source domain and target domain containing REWNs and EWPNs are first extracted from original vibration signals. The wavelet packet node energy features obtained by Daubechies2 (db2) wavelet packet decomposition were discovered to attain better recognition performance for bearing fault diagnosis after a lot of experiments on a serials of Daubechies wavelets.

Herein, the db2 is adopted as the mother wavelet function to implement binary WPD for vibration signals, and the maximum decomposition level is set to 4. After the construction of feature set, a transfer learning strategy of improving predictive performance is put forward. Additionally, to further investigate the effect of the instance weights of source-domain labeled samples and learned Mahalanobis distance metric to the overall performance, a reduction of DMTL denoted DMTL \_w only adopts instance weights, and a reduction of DMTL denoted DMTL \_l only considers the learned Mahalanobis distance metric. Where the metric \( \mathbf{A} \) is fixed to an identity matrix for DMTL \_w, and the instance weights are fixed to one for DMTL \_l.

To validate the advantage of the proposed DMTL method, several popular state-of-art supervised learning methods and transfer learning methods are conducted for comparison, including Support Vector Machine (SVM), k-nearest neighbor (KNN), Transfer Component Analysis (TCA), Geodesic Flow Kernel (GFK), Deep neural network for domain Adaptation in Fault Diagnosis (DAFD), and TrAdaBoost. SVM and KNN are conventional pattern recognition methods. TCA, GFK, and DAFD are subspace-based transfer learning methods. TCA is the representative method of feature-based transfer learning approaches, GFK learns the transferable features by constructing geodesic flow kernel, DAFD is a transfer learning method based on deep neural network which has been widely used in fault diagnosis. TrAdaBoost is the representative method of instance-based transfer learning approaches. Detailed parameter settings of above methods are described as follows. The Guassian kernel is applied in SVM classifier, and the tradeoff parameter is 1. For KNN, the number of nearest neighbor ranges from 1 to 63, and the optimal results are selected. For TCA, the optimal hyperparameters are obtained by Bayesian optimization approach, where the regularization tradeoff parameter \( \mu \) ranges from \( 10^{-1} \) to \( 10^1 \), and the subspace dimension \( l_1 \) ranges from 1 to 10. For GFK, the subspace dimension ranges from 1 to 5. For DAFD, three layers neural network are employed to extract sensitive features, and the SVM classifier is applied to identify different fault type. For TrAdaBoost, the linear SVM are adopted for base classifier, and the maximum number of iterations is set to 100. For the proposed DMTL method, tradeoff parameters \( \lambda \) and \( \beta \) are selected in the range \( [10^{-1}, 10^3] \), and the maximum number of iterations is set to 100. In all the cross-domain adaptation tasks, the labeled samples in source domain and about 15% labeled samples in target domain are adopted for training, and the remaining samples in target domain are selected for testing. The evaluation metric is the classification accuracy on the testing samples in target domain, \( \text{Acc} = \frac{1}{M} \sum_{i=1}^{M} \delta (y_i, \hat{y}_i) \), which is widely employed in literatures.

1. **Results for bearing fault diagnosis:** The classification results of bearing fault diagnosis tasks are shown in Table 2. The average accuracies of SVM, KNN, TCA, GFK, DAFD, TrAdaBoost, DMTL \_w, DMTL \_l and DMTL method are 71.1%, 74.2%, 53.9%, 71.7%, 43.5%, 67.7%, 67.9%, 76.0%, and 80.5%, respectively. As seen, none of the comparison methods achieves the best classification results on all the tasks. DMTL outperforms than other compared approaches by a significant margin in cross-domain adaptation scenarios for A—C and B—A, which attains the highest average transfer recognition accuracies of 80.5%. The recognition performance of

| Tasks | SVM | KNN | TCA | GFK | DAFD | TrAdaBoost | DMTL \_w | DMTL \_l | DMTL |
|-------|-----|-----|-----|-----|-----|------------|--------|--------|------|
| A—B   | 85.3| 85.9| 68.8| 90.0| 46.9 ± 10.9 | 85.3 | 75.3 | 85.6 | 86.6 |
| A—C   | 80.4| 68.6| 33.3| 69.0| 37.5 ± 7.2  | 48.1 | 61.2 | 67.8 | 80.6 |
| B—A   | 85.6| 88.5| 57.6| 81.2| 59.8 ± 3.9  | 72.8 | 70.8 | 91.2 | 91.5 |
| B—C   | 87.3| 70.2| 33.3| 69.4| 39.3 ± 8.7  | 50.4 | 66.7 | 72.8 | 82.4 |
| C—A   | 55.0| 61.5| 70.0| 60.3| 40.3 ± 6.3  | 74.3 | 61.8 | 68.9 | 69.1 |
| C—B   | 33.3| 70.3| 60.3| 60.0| 37.4 ± 6.8  | 75.0 | 63.4 | 69.8 | 72.6 |
| Average| 71.1| 74.2| 53.9| 71.7| 43.5         | 67.7 | 67.9 | 76.0 | 80.5 |

**Table 2.** Classification accuracy (%) of domain adaptation results in bearing fault diagnosis.
DMTL \_l and DMTL is better than DMTL \_w. Since the metric $A$ is replaced by an identity matrix in DMTL \_w, the instance weights are ineffective. While in DMTL \_l, the instance weights of source domain and target domain are equal, which fails to bridge the gap between domains. The results indicate that the proposed DMTL method is able to effectively recognize different rolling bearing conditions even there is only a small amount of labeled data. In the adaptation tasks $A \rightarrow B$ and $B \rightarrow A$, almost all methods achieved high accuracy, because datasets $A$ and $B$ are collected from the same platform with small difference in distribution, which implies that all algorithms have better transfer performance of datasets collected under different working conditions. While the transfer performance may decline when the dataset C is used for source domain, which is because datasets A and B contain more operating condition types than dataset C. Datasets C collects three operating conditions (N, OR, IR) while datasets A and B includes one more operating conditions: ball faults (B). The lack of knowledge in ball faults may decline the transfer learning performance and result in lower diagnosis accuracy when using dataset C as the source domain, and introducing a small number of labeled training samples in target domain can improve the transfer performance. More representative features that are shared among with source domain and target domain can be effectively extracted for a larger domain shift by DMTL, which can be better analyzed subsequently. Thus, DMTL is more effective when the domain shift is larger, which makes DMTL suitable for complex adaptation tasks and real-world applications.

(2) Results for Real-World Run-to-Failure Stage Diagnosis: The results for diagnosing fault severity of bearings are presented in Table 3. The average accuracies of SVM, KNN, TCA, GFK, DAFD, TrAdaBoost, and DMTL method are 86.7\%, 93.4\%, 52.4\%, 85.9\%, 47.8\%, 94.8\%, and 98.1\%, respectively. Obviously, the proposed DMTL algorithm achieves the best diagnostic performance with recognition accuracy of 98.1\% in all compared methods, while DAFD provides the worst average diagnostic accuracy at 47.8\% because the lack of training samples for transfer learning based on deep neural network. TrAdaBoost and KNN gain second-best average diagnostic accuracy at 94.8\% and 93.4\%, respectively. Almost all the methods achieve worst transfer performance shown in the first three rows of Table 3, which are the most complex cross-domain adaptation scenarios. Although the type of tested bearings is same, the severities of faults and testing environments are different, which indicates that the domain shift between the source domain and the target domain is large. Nevertheless, DMTL achieves stable diagnostic accuracy than the compared algorithms under these complex cross-domain adaptation scenarios by weakening the weight of low-quality instances and enhancing the weight of high-quality samples of source domain in training stage, which the intrinsic similarity or dissimilarity between instances are learned by Mahalanobis distance. The superior transfer performance of DMTL on slight fault, moderate fault and serious fault diagnosis suggests its capability in handling the complex cross-domain adaptation scenarios. Moreover, the ability in diagnosing early failures is helpful for early warnings in real-world applications thereby avoiding further losses.

Sensitivity analysis for different parameters in dmtl

Different parameters may influence the classification performance of DMTL, the sensitivity studies on the parameters of DMTL were conducted in this section.

### Table 3. Classification accuracy (%) of domain adaptation results in bearing fault severity diagnosis.

| Tasks       | SVM  | KNN  | TCA  | GFK  | DAFD | TrAdaBoost | DMTL |
|-------------|------|------|------|------|------|------------|------|
| A $\rightarrow$ slight | 93.3 | 91   | 33.3 | 72.5 | 38.4 $\pm$ 10.0 | 96.7 | 97.3 |
| B $\rightarrow$ slight | 92.5 | 94.9 | 33.3 | 78   | 36.1 $\pm$ 7.9  | 97.1 | 97.3 |
| C $\rightarrow$ slight | 49.4 | 88.2 | 89.4 | 74.1 | 56.9 $\pm$ 10.9 | 96.8 | 97.3 |
| A $\rightarrow$ moderate | 85.1 | 88.6 | 33.3 | 81.2 | 34.9 $\pm$ 5.3  | 90.3 | 98.0 |
| B $\rightarrow$ moderate | 85.9 | 86.7 | 33.3 | 81.2 | 36.6 $\pm$ 9.6  | 82.6 | 98.4 |
| C $\rightarrow$ moderate | 74.1 | 91.4 | 85.1 | 85.9 | 67.2 $\pm$ 9.2  | 89.9 | 94.5 |
| A $\rightarrow$ serious | 99.6 | 100  | 33.3 | 100  | 40.7 $\pm$ 12.7 | 100  | 100  |
| B $\rightarrow$ serious | 100  | 100  | 33.3 | 100  | 43.2 $\pm$ 13.4 | 100  | 100  |
| C $\rightarrow$ serious | 100  | 100  | 97.3 | 100  | 75.9 $\pm$ 7.4  | 100  | 100  |
| Average     | 86.7 | 93.4 | 52.4 | 85.9 | 47.8          | 94.8 | 98.1 |
Sensitivity analysis on K for KNN classifiers: In DMTL, the KNN classifier is adopted to recognize fault types by using adjusted samples of new subspace. The transfer learning performance may be affected by different numbers of nearest neighbors, and the potential impacts imposed by different K to the accuracy of DMTL on bearing fault severity diagnosis are shown in Figure 3. As seen, the accuracies of DMTL on the tasks A→serious, B→serious and C→serious are higher than that of other tasks with varying values of K. The accuracies fluctuate in all transfer tasks, there is an overall downward trend with the increasing values of K on the tasks A→light, B→light and C→light, while there is a slight increase in recognition accuracy with the increasing values of K on the tasks A→moderate, B→moderate and C→moderate. From Figure 3, a large value of K may decline the accuracy of cross-domain recognition performance, while a small value of K can obtain more stable and precise accuracy. Thus, the number of nearest neighbors K is set to a small integer, for example, $K \leq 3$.

Sensitivity analysis on number of instance pairs C: The in-class and out-of-class instance pairs C are used to estimate the loss function as shown in Eq. (9), which is randomly selected for optimization in recognition task. The impacts of different values of C to classification accuracy of DMTL in bearing fault severity diagnosis tasks are investigated in this section. Classification accuracies with varying values of C range from 50 to 1500 are displayed in Figure 4. Here, the number of nearest neighbors K is set to 3. As seen, the accuracies of DMTL on the nine transfer tasks vary slightly with increasing values of C. The accuracies of DMTL on the tasks A→serious, B→serious and C→serious are higher than that of other tasks with varying values of C, and the recognition rate is stable at 100%. There is a slight
fluctuation in recognition accuracy with the increasing values of $C$ in most of the transfer tasks, and the accuracies decrease with the increasing values of $C$ on the tasks when $C > 250$. Meanwhile, the computational cost will raise with increasing values of $C$.

(3) **Sensitivity analysis on trade-off parameters $\lambda$ and $\beta$:** The trade-off parameters $\lambda$ and $\beta$ is used to balance impacts of loss function of prediction model and instance weights terms as shown in the objective function of DMTL. The sensitivity analysis on those two parameters in B→moderate and C→moderate tasks are conducted in this section, and the experimental results are exhibited in Figure 5. Here, the number of nearest neighbors $K$ and the number of instance pairs $C$ are set to 3 and 150, respectively. Obviously, the transform performance decrease sharply when parameters $\beta > 100$, and DMTL performs well and stably when parameters $\beta$ is set to be a relatively small value, for example $\beta \leq 1$. The parameters $\lambda$ has little effect on the recognition performance, which can be set into a wide range $[10^{-3}, 10^3]$.

![Figure 5. Sensitivity analysis on trade-off parameters $\lambda$ and $\beta$.](image)

**Conclusions**

In this study, a cross-domain fault diagnosis model based on distance metric transfer learning (DMTL) is proposed to recognition the operating condition of rolling bearing when the labeled samples in target domain is insufficient. DMTL reweights samples in source domain by maximizing the intra-class distances and minimizing the inter-class distances for target domain, and the objective function is defined on basis of Mahalanobis distance instead of Euclidean distance. The features of source domain and target domain are first extracted from original vibration signals by using wavelet packet decomposition (WPD). Then, the DMTL model is adopted to eliminate the error propagation across different components, which can weaken the weight of low-quality instances and enhance the weight of high-quality samples. Finally, the k-nearest neighbor (KNN) classifier is applied to accomplish the cross-domain intelligent fault-type classification. The effectiveness and superiority of proposed DMTL and WPD method is verified through two transfer recognition experiments. Compared with other peer methods, the proposed method has better fault diagnosis effect in cross-domain adaption tasks, which implies that the proposed method possesses accurate recognition performance in target domain than other compared ones by using only a few of labeled target samples and massive source samples.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is sponsored by the National Natural Science Foundation of China (Grant No. 52005168), the Green Industry Technology Leading Program of Hubei University of Technology (XJ2021005001), and the Scientific Research Foundation for High-level Talents of Hubei University of Technology (GCRC2020009).
11. Tayyab SM, Chatterton S and Pennacchi P. Intelligent fault diagnosis of rolling element bearings under constant/variable load and speed conditions. *Mech Syst Signal Process* 2021; 149: 107153.

2. Huo C, Jiang Q, Shen Y, et al. New transfer learning fault diagnosis method of rolling bearing based on ADC-CNN and LATL under variable conditions. *Measurement* 2022; 188: 110387.

3. Zhou H, Zhong F, Shi T, et al. Class-information-incorporated kernel entropy component analysis with application to bearing fault diagnosis. *J Vib Control* 2021; 27: 543–555.

4. Pang S, Yang X, Zhang X, et al. Fault diagnosis of rotating machinery with ensemble kernel extreme learning machine based on fused multi-domain features. *ISA Trans* 2020; 98: 320–337.

5. Chen J, Wang J, Zhu J, et al. Unsupervised cross-domain feature learning using improved convolutional deep belief networks for rotating machinery. *IEEE/ASME Trans Mechatron* 2021; 26: 2770–2781.

6. Zhou H, Shi T, Liao G, et al. Weighted kernel entropy component analysis for fault diagnosis of rolling bearings. *Sensors* 2017; 17: 625.

7. Cerrada M, Sánchez RV, Li C, et al. A review on data-driven fault severity assessment in rolling bearings. *Mech Syst Signal Process* 2018; 99: 169–196.

8. Liu R, Yang B, Zio E, et al. Artificial intelligence for fault diagnosis of rotating machinery: a review. *Mech Syst Signal Process* 2018; 108: 33–47.

9. Duan J, Zhang X and Shi T. A hybrid attention-based paralleled deep learning model for tool wear prediction. *Expert Syst Appl* 2023; 211: 118548.

10. Shao H, Jiang H, Zhang H, et al. Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing. *Mech Syst Signal Process* 2018; 100: 743–765.

11. Tayyab SM, Chatterton S and Pennacchi P. Intelligent defect diagnosis of rolling element bearings under variable operating conditions using convolutional neural network and order maps. *Sensors* 2022; 22: 2026.

12. Zhao J and Huang W. Transfer learning method for rolling bearing fault diagnosis under different working conditions based on CycleGAN. *Meas Sci Technol* 2022; 33: 025003.

13. Shao L, Zhu F and Li X. Transfer learning for visual categorization: a survey. *IEEE Trans Neural Netw Learn Syst* 2015; 26: 1019–1034.

14. Yan R, Shen F, Sun C, et al. Knowledge transfer for rotary machine fault diagnosis. *IEEE Sens J* 2020; 20: 8374–8393.

15. Zhang W and Li X. Federated transfer learning for intelligent fault diagnostics using deep adversarial networks with data privacy. *IEEE/ASME Trans Mechatron* 2022; 27: 430–439.

16. Zhang W, Li X, Ma H, et al. Universal domain adaptation in fault diagnostics with hybrid weighted deep adversarial learning. *IEEE Trans Ind Inform* 2021; 17: 7957–7967.