Article

Research on Shaft Current Damage Identification of Variable Condition Motor Bearings Based on Multiscale Feature Label Propagation and Manifold Metric Transfer

Guangbin Wang [1,2], Shubiao Zhao [1,2,*], Zhixian Zhong [2] and Dong Zeng [3]

[1] School of Mechanical and Electrical Engineering, Lingnan Normal University, Zhanjiang 524048, China
[2] School of Mechanical and Control Engineering, Guilin University of Technology, Guilin 541004, China
[3] Guangdong Yuedian Zhanjiang Wind Power Co., Ltd., Zhanjiang 524048, China
[*] Correspondence: zhaonne@163.com

Abstract: The current damage is the most stubborn and difficult fault of high-power motor bearings because its vibration characteristics are easily confused with those of ordinary bearing mechanical faults. If it is discriminated as an ordinary mechanical fault without electrical insulation protection, the current damage of bearing shafts will still repeatedly appear. Aiming at the problem that it is difficult to identify the bearing current damage fault under variable working conditions, a bearing shaft current damage identification method based on multiscale feature label propagation and manifold metric transfer (MFLP-MMT) is proposed. Firstly, the multiscale sub-band signal is obtained by wavelet packet decomposition, and the multiscale sub-band fuzzy entropy is obtained by calculating its fuzzy entropy. Then, according to the extracted features, a neighbor graph is constructed on the source domain of the known fault label to obtain the pseudo label of the target domain sample, and the source domain label information is gradually diffused by way of the graph label propagation. The multiscale sub-band fuzzy entropy of the sample is mapped to the low-dimensional manifold space by locality preserving projections (LPP), and the source domain samples close to the target domain are given higher weights by cross-domain density ratio estimation to solve the problem of domain offset. Combined with the label samples of the target domain in label propagation, the manifold distance metric is learned to minimize the intra-class distance and maximize the inter-class distance in the domain and eliminate the overlapping phenomenon in the domain. By increasing the range of label propagation after each iteration, the label propagation error of the leading graph is gradually reduced, and unsupervised metric transfer learning is realized. The experimental results show that the new method is superior to the semi-supervised transfer learning method in fault identification ability; the highest fault identification accuracy can reach 100% and it has a good robustness.

Keywords: current damage; transfer learning; multiscale feature; label propagation; manifold space

1. Introduction

With the wide application of pulse width modulation inverters in motors, bearing current damage has become an important concern for high-power motors. When the motor is running, due to the fast change of the frequency of the variable frequency drive and the high pulse voltage, it is easy to generate the threshold voltage of the lubricating oil film at both ends of the shaft, which exceeds the threshold voltage of the bearing, and the tip discharge occurs on the contact surface between the bearing raceway and the rolling element, causing the shaft current damage and causing safety accidents. In high-end equipment, such as high-speed rail trains, large wind turbines, and new energy buses, the power of the generators and motors is increasing, the frequency conversion control system is becoming more and more complex, and the bearing current damage is increasing, resulting in a sharp reduction in the life of the batch bearings relative to the design life. At present, the monitored bearing fault signal is often misjudged as ordinary fatigue

wear or other mechanical damage. Simply replacing the bearing without taking electrical insulation maintenance measures is palliative. If the fault type of the motor bearing cannot be accurately identified, the same fault type will repeatedly appear, delaying the best maintenance time of the motor system, causing huge safety hazards and economic losses. Therefore, the accurate identification of motor bearing faults is crucial [1].

At present, there are much research on the current damage of motor bearings, including the mechanism and key influencing parameters of the bearing current, the circuit model of the bearing current and the damage to the bearing performance [2,3]. Chen et al. [4] proposed corresponding improvement measures for the wind turbine mechanism in view of the problem of shaft voltage and shaft current of the wind turbine. Xu et al. [5] analyzed the mechanism of shaft current generation and the method of shaft current suppression in a hydro-generator. Adabi et al. [6] analyzed the causes of the shaft voltage of induction generator and proposed the suppression measures of the shaft current. Some scholars have studied the generation mechanism and damage characteristics of shaft voltage and shaft current from an electrical point of view for specific generators and motors. For example, Prashad et al. [7] established a charge accumulation and discharge energy model with the time period, corrosion pit size, and bearing capacitance as the parameters, so as to determine the shaft voltage and diagnose the bearing failure caused by corrosion pits. Liu et al. [8] analyzed the electrical discharge machining, wrinkling, and its characterization caused by bearing current from the perspective of tribology. Picot assumed that the premature failure of the bearing was caused by the high-frequency current passing through the bearing. A Weibull cycle statistical method based on the stator current was proposed to construct a stable, normalized fault index [9]. Liu et al. [10,11] established an equivalent electrical model of the bearing, simulated the process of the bearing breakdown and recovery, and proposed measures to suppress the bearing current of small generators and motors. There are also scholars from the statistical point of view using current monitoring data and vibration monitoring data to establish the identification model of bearing current damage and detect fault in bearing current damage. For example, Kempski proposed an electrostatic discharge statistical model of induction motor bearings generated by inverters, which can be used as a research basis for bearing damage risk of different drive devices [12]. Houssin proposed a stator current parameter spectrum estimation method to detect induction motor bearing faults. The effectiveness of bearing fault detection is verified by simulation analysis [13]. Based on a collaborative filtering algorithm and manifold learning algorithm, Wang et al. [14,15] studied the fault identification and state prediction of bearing current damage. Chen et al. [16] proposed a lifting wavelet fractal strategy to extract the impact fault features reflecting bearing faults from multi-component coupled vibration signals, which has a more comprehensive ability to analyze the current damage characteristics of motor bearings. Through the above literature review, it is clear that the problem of motor bearing current damage is common in high-power motors. The research on current damage mechanism, current suppression method, and current fault detection have always been a hot spot. However, most of the current research on the current damage of motor bearings summarize and analyze the causes of bearing current from the results of shaft current damage or establish a corresponding equivalent physical model to simulate the process of bearing damage. These research methods have high diagnostic accuracy for motor bearing current damage, but their applicability is not wide, and the cost is high. There are also some studies based on data, but in the study of bearing current damage, bearing current damage is greatly affected by factors such as motor speed [17–19], load [20], and grease [21,22]. Different working conditions will lead to different data distribution, thus greatly reducing the efficiency of the fault diagnosis [23].

The traditional data-driven rolling bearing fault diagnosis method requires two basic premises: the distribution of the training data and test data is the same; there are enough available training samples [24]. In view of the need for a large number of samples, many studies have effectively alleviated the problem of insufficient numbers of labeled samples through the construction of semi-supervised learning frameworks [25,26]. The same dis-

tribution of data requires the same environmental conditions when collecting the bearing vibration signals. In the context of the Internet of Things and big data, most machine learning will collect data in various parts of the machine that may fail [27] and will train multiple fault diagnosis models. This method can also solve the problem of different data distributions, but it will consume huge manpower and material resources and is expensive. The transfer learning emerges as the time requires. The transfer of existing knowledge solves the learning problems that are difficult to obtain for the label samples in the target domain. The method of applying the knowledge learned in a domain (source domain) to different but related domains (target domain) is called transfer learning [28]. Transfer learning also has many applications in the field of bearing fault diagnosis [29–31]. Considering that the bearing vibration data under variable working conditions will have similar fault characteristics, minimizing the cross-domain distribution distance has become a hot issue to be solved in the field of bearing fault diagnosis. Lan et al. [32] proposed a cross-domain bearing fault classification method based on transfer component analysis (TCA), which significantly reduces the inter-domain distribution distance by minimizing the maximum mean difference between the source domain and the target domain and achieves cross-domain edge distribution alignment. The manifold embedded distribution alignment (MEDA) method is a transfer manifold learning method with dynamic distribution adaptability, proposed by Wang et al. [33] on the basis of joint distribution adaptation (JDA) [34], which achieves better domain adaptation by quantitatively evaluating the relative importance of the edge distribution and conditional distribution. Zhang et al. [35] proposed a small sample bearing fault diagnosis method based on transfer learning, using a sufficient number of the source domain samples to train the network to prevent network overfitting, and using 1% of the target domain training set data to fine-tune the model classification ability. Zhao et al. [36] used bidirectional gated recurrent units to generate auxiliary samples for the source domain of MEDA, so that excellent fault identification can be maintained in the case of a small number of labeled samples.

However, the above research focuses on the alignment of the edge distribution and the conditional distribution. As a cross-domain alignment, these two distribution alignments ignore the internal structure information of the domain, and the shaft current damage characteristics of the motor bearing and the common fault characteristics are easily confused, resulting in category overlap in the domain. In view of the internal structure of the field, many scholars have applied metric learning to better learn the distance or similarity between samples. By changing the original sample distribution, the intra-class dispersion and inter-class similarity are reduced, and the recognition ability of the classifier is improved [37]. Xu et al. [38] combined transfer learning with metric learning to improve the accuracy of the fault feature clustering by minimizing the correlation alignment loss, narrowing the distribution difference between the source domain and the target domain, and maximizing the similarity between the input feature and the center feature. Zhao et al. [39] used triplet loss to measure the distance between various faults, so that the distance between similar fault features is very small, and the distance between heterogeneous fault features is very large. However, the above deep metric transfer learning applied to bearing fault diagnosis focuses on the Euclidean distance between the classes in the alienation domain, which leads to the neglect of spatial cross-domain tilt and manifold structure mining. As the acquisition of the equipment status information becomes more and more difficult, the amount of fault data is small, and it is impossible to effectively label all working condition data, so the label samples are very precious [40]. In order to make efficient use of the existing fault data and fully exploit the manifold similarity between data under the diversity, time-varying, and strong nonlinearity of the working conditions, it is very important to pay attention to cross-domain tilt and intra-domain alienation.

In view of the extreme service environment and complex working load of motor bearing, the damage state signal of motor bearing presents the characteristics of weak and nonlinear and is often submerged in strong background noise and interference signal. Its common features are difficult to fully characterize as the shaft current damage of motor

bearing, and the recognition effect under variable working conditions is poor. A method of shaft current damage identification of motor bearing under variable working conditions based on multiscale feature label propagation and manifold metric transfer is proposed. Firstly, the sensitive feature set of bearing current damage is constructed by extracting the multiscale sub-band fuzzy entropy features of the signal, which can better identify the bearing current damage. Then, according to the proposed features, the weighted graph is constructed on the source domain of the known fault label, and the source domain label information is gradually diffused by graph label propagation, which solves the limitation that the semi-supervised metric transfer learning needs supervised samples. At the same time, the LPP algorithm is used to map the multiscale sub-band fuzzy entropy of the sample to the low-dimensional manifold space, and the manifold distance metric is learned to align the source domain and the target domain. The intra-class distance is minimized, and the inter-class distance is maximized to eliminate the overlap in the domain and realize unsupervised metric transfer learning. In the case of no target domain label samples, the problem of poor bearing current damage identification under different working conditions is solved. The proposed method mainly has the following four advantages: (1) This method belongs to the unsupervised transfer learning method, which breaks the limitation that the traditional metric transfer learning needs the target domain sample label. (2) This method considers the manifold relationship between data samples, mining the manifold structure between data. (3) The proposed method takes into account the problems of domain tilt, excessive inter-domain distance, and intra-domain overlap, which are neglected in the current damage identification of motor bearing shaft under variable working conditions. (4) The multiscale sub-band fuzzy entropy extracted by this method can dig deep into the essential information of the vibration information and can accurately identify the current damage fault of the motor bearing.

2. Multiscale Sub-Band Fuzzy Entropy and Metric Transfer Theory

2.1. Multiscale Sub-Band Fuzzy Entropy

Fuzzy entropy is a measure of the complexity of time series. It describes the information contained in time series samples more accurately by introducing the fuzzy membership function. However, when the sequence is too complex, the ordinary fuzzy entropy extraction cannot dig deep into the internal information of the data. In order to solve this problem, we put forward the concept of multiscale sub-band fuzzy entropy. Firstly, the signal is decomposed into multiscale sub-band signals by wavelet packet, and, finally, the fuzzy entropy of each multiscale sub-band signal is obtained. Because this method can go deep into the data and decompose and mine the eigenvalues and eigenvectors of the data layer by layer, it can effectively and accurately extract the fault features of complex weak data. The specific steps of multiscale sub-band fuzzy entropy are:

(1) The vibration signal $X(t)$ is decomposed and reconstructed by N layer wavelet packets, and 2^N wavelet packet decompositions and reconstruction sequences $S(N, k)$ $(k = 0, 1, 2, \cdots, 2^N - 1)$ are obtained. $S(N, k)$ is the kth sub-band sequence of the N layer decomposition of signal $X(t)$ by the wavelet packet.

(2) Taking the sub-band $S(N, 0)$ as an example, given the dimension $m = 2$, the sub-band signal $S(N, 0)$ is composed of a set of m dimensional vectors, according to the serial number, namely:

$$S(N, 0)(i) = [S(N, 0)(i), \cdots, S(N, 0)(i + m - 1)] \tag{1}$$

where $i = 1, 2, 3, \cdots, B - m + 1$ and B is the length of the sub-band sequence.

(3) The distance is calculated between each sequence and the remaining sequences d_{ij}:

$$d_{ij} = d[S(N, 0)(j) - S(N, 0)(i)] = \max \left| S(N, 0)(j + l) - S(N, 0)(i + l) \right| \tag{2}$$

where $l = 0, 1, 2, \cdots, m - 1$

(4) Given a threshold $r = 0.2S_{td}$ (S_{td} is the standard deviation of the original data), ambiguity $\tau = 2$, through the fuzzy membership function $\mu\left(d_{ij}^m, \tau, r\right)$, the distance matrix is redefined:

$$D_{ij}^m = \mu\left(d_{ij}^m, \tau, r\right) = \exp\left(\frac{-\left(d_{ij}^m\right)^\tau}{r}\right) \tag{3}$$

(5) The average of all memberships is found:

$$\phi^m(\tau, r) = \frac{1}{B-m} \sum_{i=1}^{B-m} \left(\frac{1}{B-m-1} \sum_{j=1, j\neq i}^{B-m} D_{ij}^m\right) \tag{4}$$

(6) The dimension is increased to $m + 1$, steps (2)–(5) are repeated to obtain ϕ^{m+1}.

(7) The fuzzy entropy of $S(N, 0)$ is:

$$FuzzyEn(m, \tau, r, B) = \ln \phi^m(\tau, r) - \ln \phi^{m+1}(\tau, r) \tag{5}$$

(8) Steps (2)–(7) are repeated and the fuzzy entropy of sub-band $S(N, 0), S(N, 1)$, $S(N, 2), \cdots, S(N, k)$, respectively, is calculated.

(9) In this paper, the sub-band sequences $S(N, k)(N = 5)$ $k = 0, 1, 2, \cdots, 31$ are calculated, respectively, and the characteristic matrix is constructed. The multiscale sub-band fuzzy entropy of the fault signal is F:

$$F = [c(5, 0), c(5, 1), c(5, 2), c(5, 3)\ldots\ldots c(5, 31)] \tag{6}$$

2.2. Metric Transfer Learning

Metric learning can adaptively learn metrics from raw data with the same distribution, change the distribution of the original samples, reduce the distance between similar samples, and dissimilate heterogeneous samples to greatly improve the efficiency of machine learning. Metric transfer learning reduces the difference of the sample distribution caused by different data distributions and maintains better classification ability when data are distributed differently.

For a given sample set $X = [x_1, x_2, \ldots x_n] \in \mathbb{R}^{n\times D}$, the core goal of metric learning is to find an optimal learning metric A, under which the samples of the same class are more aggregated in space and alienated into different classes. Therefore, Jin et al. formulated the following regular objective functions [41]:

$$\min_{A} \left\{ \frac{1}{2} \parallel A \parallel_F^2 + \frac{2}{n(n-1)} \sum_{i<j} \ell\left(\delta_{ij}\left[1 - \parallel x_i - x_j \parallel_A^2\right]\right) \right\} \tag{7}$$

where A is a positive semidefinite matrix; $\ell(\cdot)$ denotes the loss function; δ_{ij} is the indicator matrix within and between classes. When two samples are of the same class, $\delta_{ij} = 1$, otherwise $\delta_{ij} = -1$.

According to the above objective function, a distance metric A can be obtained. Under this metric, the distance between sample x_i and sample x_j can be expressed as:

$$d_{ij} = \sqrt{(x_i - x_j)^T A (x_i - x_j)} \tag{8}$$

Traditional distance metric learning is based on the premise that all samples obey the same distribution. When the distribution of samples in the space is inconsistent, the effect of metric learning will decline or even fail. Therefore, after the metric transfer learning [42] introduces sample cross-domain information in the metric learning, the loss function becomes:

$$\ell(A, \hat{\omega}, D_s, D_t) = \ell_w(A, \hat{\omega}) - \ell_b(A, \hat{\omega}) \tag{9}$$

where $\hat{\omega}$ is the sample weight; $\ell_w(A, \hat{\omega})$ and $\ell_b(A, \hat{\omega})$ represent the loss functions within and between classes. The specific expressions are as follows:

$$\ell(A, \hat{\omega}) = \sum \hat{\omega}(x_i)\hat{\omega}(x_j)\delta_{ij}\|x_i - x_j\|_A^2 \tag{10}$$

Metric transfer learning solves the domain offset problem by assigning different weights to the labelled samples, and clustering and alienating samples in the field on the basis of cross-domain alignment, and maximizes the class spacing to obtain the distance metric A.

3. Label Propagation and Manifold Metric Transfer Learning Algorithm

Due to the difference of spatial distribution between the source domain and the target domain, simple feature normalization in the domain cannot solve the problems of domain offset and domain tilt. When the number of source domain samples of known labels is scarce, the effect of the source domain and the target domain alignment often decreases greatly, and the existing transfer learning methods pay more attention to the overall cross-domain alignment of the domain and less attention to the distribution within the domain. In the motor bearing shaft current damage identification, the similar features between the shaft current damage and ordinary fault leads to different types of overlap. Many machine learnings directly measure the similarity between data in the original Euclidean space, ignoring the hidden manifold structure relationships in the sample space. Therefore, in order to mine the manifold similarity between data, pay attention to the structural information between and within the domain, and get rid of the dependence on the label samples of the target working condition. An unsupervised transfer learning method based on multiscale feature label propagation and manifold metric transfer is proposed, aiming at eliminating the domain overlap and intra-domain overlap in the manifold space.

3.1. Graph-Based Label Propagation

The core idea of graph-based label propagation (GLP) is that similar samples will have similar label graphs. Label propagation propagates samples with known labels to predict unknown labels through connected edges [43]. Commonly used graphs have weighted graphs. Each sample can be regarded as a node. The connection weight of Node i and Node j can be expressed as:

$$G_{ij} = \begin{cases} \exp\left(-\dfrac{|x_i - x_j|^2}{\alpha^2}\right), & x_i \in \mathcal{N}_p(x_j) \text{ or } x_j \in \mathcal{N}_p(x_i) \\ 0, & \text{others} \end{cases} \tag{11}$$

where $\mathcal{N}_p(x_i)$ represents the p nearest neighbor set of point x_i. The loss function of cross-domain label propagation from the source domain to the target domain is defined as follows:

$$\begin{aligned} \mathcal{J}_{Lp} &= \sum_{i,j=1}^{n_s+n_t} G_{ij}\|F_i - F_j\|_F^2 = \mathrm{tr}\left(F^T(D - G)F\right) \\ &= \mathrm{tr}\left(\begin{bmatrix} F_s \\ F_t \end{bmatrix}^T \begin{bmatrix} L_{ss} & L_{st} \\ L_{ts} & L_{tt} \end{bmatrix} \begin{bmatrix} F_s \\ F_t \end{bmatrix}\right) \\ &= \mathrm{tr}\left(F_s^T L_{ss} F_s\right) + \mathrm{tr}\left(F_t^T L_{tt} F_t\right) + 2\mathrm{tr}\left(F_t^T L_{ts} F_s\right) \end{aligned} \tag{12}$$

where, F_s is the label matrix of the source domain $F_s \in R^{n_s \times C}$. F_t is the label matrix of the target domain, and the same value is initialized. $L = D - G$ represents the Laplace matrix of the graph, D is a diagonal matrix, and its diagonal elements are the sum of the column elements of the G matrix.

$$L = \begin{bmatrix} L_{ss} & L_{st} \\ L_{ts} & L_{tt} \end{bmatrix} \tag{13}$$

The label matrix F_t of the target domain can be obtained by minimizing the loss function of label propagation. Because the target domain does not provide the supervision information of labels, there are errors in cross-domain label propagation, and then we gradually reduce the label errors by iterative method.

3.2. Locality Preserving Projections

In the original space, the sample distribution is different because of the different data acquisition environment between the source domain and the target domain. Many machine learning methods use Euclidean distance to measure the similarity between samples or transfer the source domain and the target domain directly in Euclidean space, which do not consider that the nonlinear manifold structure between the samples may not achieve the ideal transfer effect. Therefore, we use the locality preserving projection in manifold learning to mine the hidden manifold structure in Euclidean space and to find the similarity between samples, which is difficult to obtain.

The LPP algorithm [44] breaks the limitation of traditional methods such as PCA, which have difficulty in mining the nonlinear manifold of data and can obtain low-dimensional projection more easily. LPP is also a graph-based manifold learning method. For the data set $X = [x_1, x_2, \ldots x_{n_s+n_t}] \in \mathbb{R}^{(n_s+n_t) \times D}$ in the original space, let $Z = [z_1, z_2, \ldots z_{n_s+n_t}] \in \mathbb{R}^{(n_s+n_t) \times d}$ be the data in the original space and map it to the data in the low-dimensional manifold. By constructing a weighted graph G of the sample, the connected points in the weighted graph remain connected after the manifold mapping, and the error is minimized as the objective function:

$$min \ \mathcal{J}_{LPP} = \sum_{i,j=1}^{n_s+n_t} \frac{1}{2} \parallel z_i - z_j \parallel^2 G_{ij} \tag{14}$$

The objective function is simplified:

$$
\begin{aligned}
min \ \mathcal{J}_{LPP} &= \sum_{i,j=1}^{n_s+n_t} \frac{1}{2} \parallel z_i - z_j \parallel^2 G_{ij} \\
&= \sum_{i,j=1}^{n_s+n_t} \frac{1}{2} \parallel R^T x_i - R^T x_j \parallel^2 G_{ij} \\
&= \text{tr}(R^T X (D - G) X^T R) = \text{tr}(R^T X L X^T R)
\end{aligned}
\tag{15}
$$

where D is a diagonal matrix whose diagonal elements are the sum of the corresponding column elements in G.

By adding constraint $R^T X L X^T R = I$, the target function is converted to:

$$\min tr\left(R^T X L X^T R\right) \ s.t. R^T X L X^T R = I \tag{16}$$

The optimization problem of the above formula is solved according to Lagrange multiplier method, which solves the eigenvalue and eigenvector problems of the generalized eigenequation of the following formula [45]:

$$X L X^T R = \lambda X D X^T R \tag{17}$$

The transformation matrix $R = [r_0, r_1, \ldots r_{l-1}] \in \mathbb{R}^{n \times l}$ is composed of eigenvectors corresponding to the first l minimum eigenvalues of the above formula.

3.3. Label Propagation and Manifold Metric Transfer Learning

The label propagation and manifold metric transfer learning aims to find a manifold distance metric that can make the same kind of samples more clustered and the different kind of samples more dispersed in the domain, with the smallest inter-domain offset by label diffusion and metric learning on the graph of low-dimensional manifolds.

3.3.1. Cross-Domain Alignment Based on GLP

Because of the differences in the spatial distribution of data samples, there will be differences in similarity between the source domain samples and the target domain samples. We first map the samples in the original space to the low-dimensional manifold space and then give higher weights to the samples similar to the source domain and the target domain, which cannot only minimize the cross-domain manifold distance but also eliminate the cross-domain tilt phenomenon. Firstly, the prior weight of all labelled samples in the source domain is assumed to be $\omega_0(z_i^s) \in \mathbb{R}^{n_s \times 1}$, and then a more accurate source domain sample weight $\omega(z_i^s)$ is obtained by introducing a regularization term to minimize the difference between the source domain and the target domain. The regularization term is expressed as:

$$\psi(\omega) = \|\omega - \omega_0\|^2 \tag{18}$$

where $\omega_0(z_i^s) = \frac{P_t(z_i^s)}{P_s(z_i^s)}$ is the density ratio estimation of the source domain sample z_i^s, which is used to express the weight of the samples in the cross-domain transfer. The higher the value of $\omega_0(z_i^s)$, the more similar the source domain sample z_i^s is to the target domain sample, and the more important it is to measure the cross-domain manifold distance.

We use the method in reference [46] to estimate the density ratio of samples in the source domain. The density ratio can be approximately considered as a combination of some basic linear functions:

$$\omega_0(z_i^s) = \sum_{j=1}^{n_t^L} \alpha_j \phi_j(z_i^s) \tag{19}$$

where ϕ_j represents a predefined Gaussian kernel function; α_j is the parameter corresponding to the basis function. Because different basis function ϕ_j settings will affect the result of the density ratio estimation, sample c_j is selected from the target domain as the center of the Gaussian function (these samples are selected by GLP). The basis function ϕ_j of Gaussian kernel is:

$$\phi_j(z) = \exp\left\{ -\frac{\|z - c_j\|^2}{\sigma^2} \right\} \tag{20}$$

Therefore, the weight $\omega_0(z_i^s)$ of the source domain samples can be solved by minimizing the KL-divergence between $P_t(z_i^s)$ and $\omega_0(z_i^s) P_s(z_i^s)$.

According to the above, it can be summarized as the following optimization objectives:

$$\max_{\alpha} \sum_{z_i \in D_t} \log \sum_{i=1}^{n_t^L} \alpha_j \phi_j(z_i^s) \quad \text{s.t.} \sum_{z_i \in D_s} \sum_{j=1}^{n_i^l} \alpha_j \phi_j(z_i^s) = n_s, \text{and} \alpha > 0 \tag{21}$$

For the above convex optimization problem, we can obtain the optimal solution by the gradient convergence method.

In the above formula, the cross-domain transfer weights $\omega_0(z_i^s)$ of all samples in the source domain are obtained. In order to make the source domain samples with high weight as close to the target domain as possible, semi-supervised metric transfer learning usually requires providing some supervised samples in the target domain. We select the supervised samples based on the GLP method, estimate the reconstruction density ratio ω_0 to $\hat{\omega}_0$, and give the target domain samples a weight of 100%:

$$\hat{\omega}_0(z_i) = \begin{cases} \omega_0(z_i), & z_i \in D_s \\ 1, & z_i \in D_t^{pl} \end{cases} \tag{22}$$

where D_t^{pl} denotes a subset of samples learned according to the auxiliary metrics selected in the target domain, whose labels are pseudo labels.

3.3.2. Optimization Goal of Intra-Domain Alienation

In order to fully explore the similarity between the sample data in manifold space and LPP, a manifold mapping method is used based on the graph. Therefore, we introduce a regularization term $\Phi(G)$ of graph:

$$\Phi(G) = \sum_{i,j=1}^{n_s+n_t} G_{ij} \parallel z_i - z_j \parallel_A^2 = \sum_{j=1}^{n_s+n_t} G_{ij}(z_i - z_j)^T A(z_i - z_j) = 2\text{tr}\left(AZLZ^T\right) \tag{23}$$

where $L = D - G$, D is the diagonal matrix of the graph matrix G.

Combined with metric learning, we can obtain the following overall optimization goals:

$$\begin{aligned}
\min_{A,\hat{\omega}} \mathcal{J} &= \lambda\psi(\hat{\omega}) + r(A) + \sigma\ell(A,\hat{\omega},D_s,D_t) + \gamma\Phi(G) \\
&= \lambda \parallel \hat{\omega} - \hat{\omega}_0 \parallel^2 + \tfrac{1}{2} \parallel A \parallel_F^2 + \\
&\quad \sigma(\ell_w(A,\hat{\omega}) - \ell_b(A,\hat{\omega})) + 2\gamma tr\left(AZLZ^T\right) \\
&= \lambda \parallel \hat{\omega} - \hat{\omega}_0 \parallel^2 + \tfrac{1}{2} \parallel A \parallel_F^2 + \\
&\quad \sigma\sum_{i,j}\hat{\omega}(z_i)\hat{\omega}(z_j)\delta_{ij} \parallel z_i - z_j \parallel_A^2 + 2\gamma tr\left(AZLZ^T\right)
\end{aligned} \tag{24}$$

where δ_{ij} is the indicator matrix within and between classes:

$$\delta_{ij} = \begin{cases} 1, & z_i, z_j \in c \\ -1, & z_i \in c, z_j \notin c \end{cases} \tag{25}$$

In order to solve the optimization problem of the objective function, it is necessary to converge A and $\hat{\omega}$ while satisfying constraints $\sum_{i=1}^{n_s}\hat{\omega}(z_i) = n_s$, and $\hat{\omega}(z_i) \geq 0$, so the Lagrange multiplier is introduced to transform the objective function into:

$$\begin{aligned}
\min_{A,\hat{\omega}} \mathcal{J} &= \lambda\|\omega - \omega_0\|^2 + \tfrac{1}{2} \\
&\parallel A \parallel_F^2 + \sigma\sum_{i,j}\hat{\omega}(z_i)\hat{\omega}\left(z_j\right)\delta_{ij} \parallel z_i - z_j \parallel_A^2 + 2\gamma\text{tr}\left(AZLZ^T\right) \\
&+ \rho((\hat{\omega}^T I - n_s)^2 + \sum_{i=1}^{n_s}(\max(0, -\hat{\omega}(z_i)))^2)
\end{aligned} \tag{26}$$

where I is the label field indication matrix for distinguishing the selected samples of the source domain and the target domain:

$$I = \begin{cases} 1, & i \leq n_s \\ 0, & n_s < i \leq n_s + n_t^{pL} \end{cases} \tag{27}$$

The objective function obtains partial derivatives of A and $\hat{\omega}$, respectively:

$$\frac{\partial \mathcal{J}}{\partial \hat{\omega}} = 2\lambda(\hat{\omega} - \hat{\omega}_0) + 2\sigma\sum_{i,j}\hat{\omega}(z_j)\delta_{ij}\|z_i - z_j\|_A^2 + \rho\left(2\left(\hat{\omega}^T I - n_s\right)I + \hat{\omega}^2\xi\right) \tag{28}$$

$$\frac{\partial \mathcal{J}}{\partial A} = A + 2\sigma\sum_{i,j}\hat{\omega}(z_i)\hat{\omega}(z_j)\delta_{ij}A(z_i - z_j)(z_i - z_j)^T + 2\gamma\text{tr}\left(ZLZ^T\right) \tag{29}$$

where $\xi_i = sign(max(0, -\hat{\omega}(z_i)))$, sign is a symbolic function.

The values of A and $\hat{\omega}$ are then alternately updated using a gradient descent to obtain the cross-domain manifold distance metric A when the convergence condition, Formula (32), of the formula is satisfied.

$$\hat{\omega}_{t+1} = \hat{\omega}_t - u_1 \frac{\partial \mathcal{J}}{\partial \hat{\omega}}|_{\hat{\omega}_t} \tag{30}$$

$$A_{t+1} = A_t - u_2 \frac{\partial \mathcal{J}}{\partial A}|_{A_t} \tag{31}$$

$$|\mathcal{J}(A_{t+1}, \hat{\omega}^{t+1}) - \mathcal{J}(A_t, \hat{\omega}^t)| < \varepsilon \tag{32}$$

Because the distance metric A is a positive semidefinite matrix, we can decompose $A = M^T M$. For simplicity, the distance between sample z_i and sample z_j can be expressed as:

$$d_{ij}^2 = (z_i - z_j)^T A (z_i - z_j) = (Mz_i - Mz_j)^T (Mz_i - Mz_j) \tag{33}$$

Under this manifold distance measure, it can be regarded that the samples after the LPP dimension reduction are mapped to a common subspace in the manifold space, where the sample distribution data in this space has the smallest cross-domain distance and the smallest cross-domain tilt, and the alienation between the clusters of the same class in the domain is obvious:

$$\begin{cases} Z_s^{new} = Z_s M \\ Z_t^{new} = Z_t M \end{cases} \tag{34}$$

4. Shaft Current Damage Mechanism and Vibration Signal Characteristics of Motor Bearings

During the normal operation of the high-power motor, there is a lubricating oil film between the bearing rolling element and the raceway, which plays an insulating role. Both ends of the spindle will only produce a lower shaft voltage. When there are problems in the design and adjustment of the motor, the failure of the electrical system makes the shaft voltage increase, the insulation oil film between the bearings is destroyed, or the stable lubricating oil film is not formed in the bearing when the motor is just started, the electric potential difference between the inner and outer rings of the bearing exceeds the breakdown voltage that the lubricating oil film can withstand; the shaft voltage can easily break the oil film and discharge, resulting in shaft current. Because the contact point between the bearing raceway and the rolling element metal is very small, the skin effect phenomenon is generated on the surface of the bearing raceway, so the current density of these points increases, and the high temperature is instantaneously generated, so that the local melting or gasification of the bearing is caused. The molten bearing metal particles splash on the surface of the bearing raceway due to the radial load grinding pressure, so that the electric erosion occurs on the bearing ball and the raceway, resulting in small pits and pitting. Under the combined action of the current and mechanical load, the electric erosion area of the bearing continues to expand, and a typical bearing current damage morphology of a 'washboard' shape is gradually formed, as shown in Figure 1.
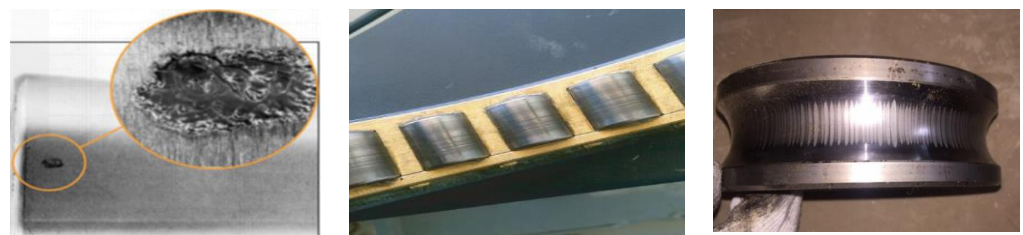

Figure 1. Corrosion spots and rubbings formed by shaft current damage.

According to the principle of shaft current damage of bearing, the bearing current damage experiment is carried out by using the bearing current damage test bench. The programmable current source intermittently loads the current of the experimental bearing, continuously loads the current of 2 A for 15 min, stops the current loading in the next 5 min, and so on. The rotational speed used in the bearing current damage experiment is 1200 rpm and the load size is constant at 1000 N. The experimental bearing is a 6205 deep-groove ball bearing with an inner diameter of 52 mm, an outer diameter of 25 mm, and with 9 balls. The characteristic frequency of the outer ring fault is fo = 71.7 Hz, the characteristic frequency of the inner ring fault is fi = 108.3 Hz, and the characteristic frequency of the rolling element fault is fb = 47.17 Hz. The time domain diagram and envelope spectrum

of the vibration signal of the shaft current damage bearing in different time periods are shown in Figures 2–4.
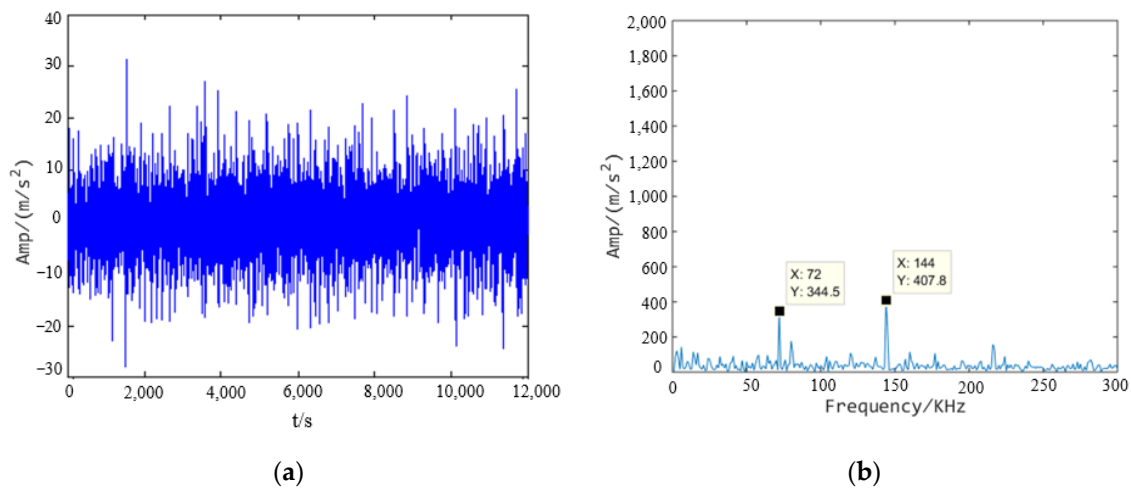

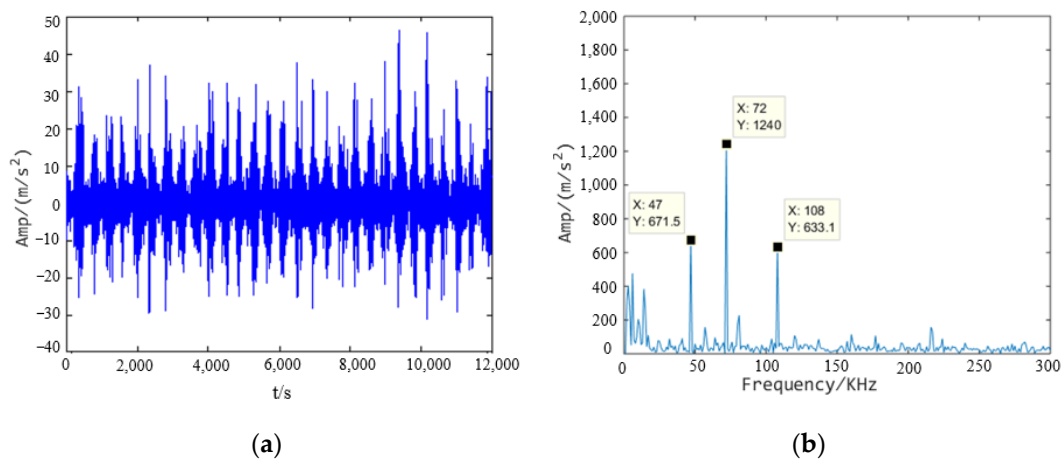**Figure 2.** Shaft current damage 40-h signal: (**a**) time domain diagram; (**b**) envelope spectrum.

Figure 3. Shaft current damage 60h signal: (**a**) time domain diagram; (**b**) envelope spectrum.
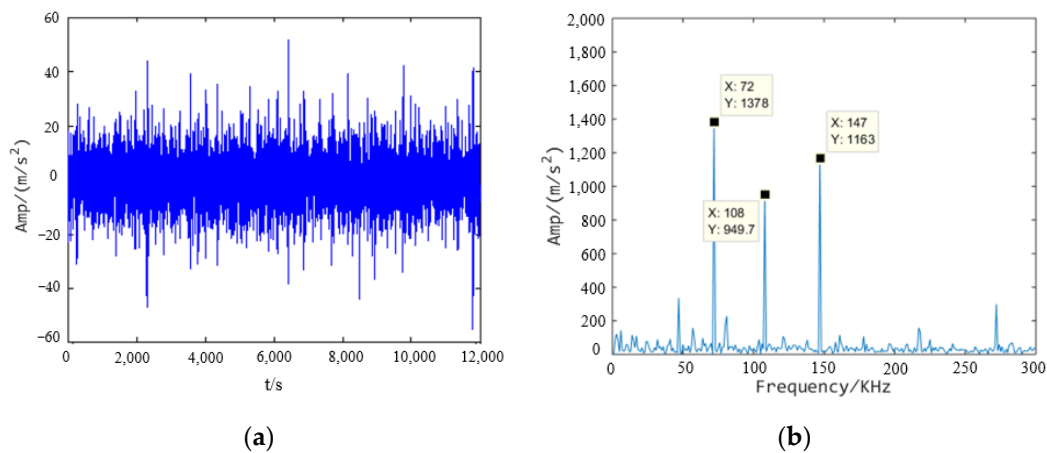

Figure 4. Shaft current damage 80-h signal: (**a**) time domain diagram; (**b**) envelope spectrum.

From Figure 2 to Figure 4, it can be seen that the fault characteristic frequency of the shaft current damage is 72 Hz at 40 h, which is characterized by the fault characteristic frequency of the outer ring of the bearing, and the double frequency amplitude is also relatively large. At 60 h, the characteristic frequencies of 47 Hz, 72 Hz, and 108 Hz can be

found in the spectrum diagram, which are the fault characteristic frequencies of the bearing balls, bearing outer rings, and bearing inner rings. In 80 h, the characteristic frequencies of 72 Hz, 147 Hz, and 108 Hz can be found in the spectrum, which are the double frequency, double frequency, and inner ring fault characteristic frequency of the outer ring fault. It can be concluded that in the shaft current damage experiment, 40 h is the initial wear stage of the bearing current damage. The initial wear pitting grooves have appeared on the bearing raceway and the ball surface, and the characteristic frequency of the outer ring fault has appeared. With the advancement of time, the bearing current damage has entered a sharp rise stage, and the fault characteristic frequency is manifested as a composite modulation form of multiple faults.

Although the bearing mechanical damage and bearing current damage bearing vibration signal characteristic frequencies in a certain correlation, because of the complexity of the bearing vibration signal of bearing current damage, only by the ordinary spectral analysis method is it difficult to effectively distinguish the bearing fault of bearing current damage and bearing mechanical damage.

5. Shaft Current Damage Identification Model of Motor Bearing Based on MFLP-MMT

Aiming at the characteristics of non-stationary, non-linear, and weak bearing fault signals, this paper proposes a fault diagnosis method combining multiscale sub-band fuzzy entropy and label propagation manifold distance metric. The multiscale sub-band fuzzy entropy can find the essential features of weak damage signals. At the same time, combined with the proposed label propagation manifold distance metric method, the distribution differences of the fault samples of motor bearings under different working conditions can be reduced, and bearing faults can be diagnosed and identified efficiently and accurately.

Based on the manifold distance metric of the graph label, the shaft current damage of motor bearing under variable working conditions is identified. Firstly, the GLP model is constructed using the samples with known fault types to obtain the pseudo labels of the unlabeled samples in the target condition. Locality preserving projection is performed on the source domain and the target domain samples and all samples are mapped to a low-dimensional manifold space. In the manifold space, by combining the labeled samples in the target domain obtained in GLP, the density ratio estimation is performed on the source domain samples to solve the sample weights, and the intra-domain metric learning is performed to obtain the optimal manifold distance metric in the manifold space. Under this metric, the samples in the source domain and the target domain can be transformed into the common subspace with the smallest cross-domain distance, the same sample clustering in the domain, and the alienation of different sample distributions, so that the differences of the sample distribution under variable working conditions can be eliminated. Finally, the fault samples are accurately classified according to the sample distribution in the manifold space.

Since this shaft current damage identification method belongs to an unsupervised transfer learning, the fault samples to be identified in the target domain are unlabeled. In the process of learning, there is a certain error in the obtained target domain label. Therefore, we will gradually reduce the identification errors of the target domain samples by increasing the label propagation range in the iteration. The overall process of motor bearing shaft current damage identification proposed in this paper is shown in Figure 5. The specific process can be divided into the following six steps:
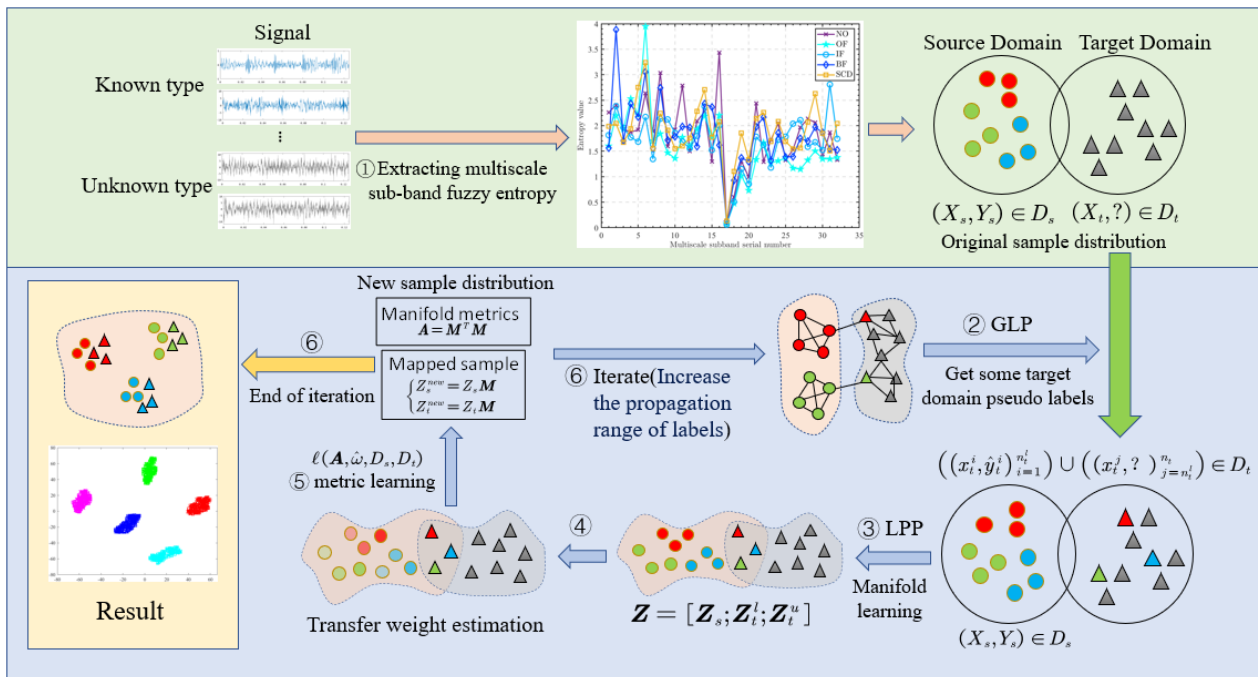
Figure 5. Overall process of shaft current damage identification for motor bearings.

Step 1: The bearing vibration signal with known label in one working condition is taken as the source domain D_s and the signal data with an unknown label in another working condition is taken as the target domain D_t. The vibration signal samples in the source domain and the target domain are extracted by multiscale sub-band fuzzy entropy as input.

Step 2: Based on GLP, transmits of the label information of the source domain samples to the target domain are used to obtain the pseudo label $\hat{y}_t$ of the target domain. Based on the result of label propagation, a small part of the sample $D_t^l = ((x_t^i, \hat{y}_t^i)_{i=1}^{n_t^l})$ with the richest label information in the target domain is retained, that is, the sample closest to the source domain on the weighted graph. This part of the sample has a relatively small label error, and we retain this part of label information to obtain a new target domain $D_t = ((x_t^i, \hat{y}_t^i)_{i=1}^{n_t^l}) \cup ((x_t^j, ?)_{j=n_t^l}^{n_t})$ with partial labels.

Step 3: According to the source domain sample D_s and the target domain sample D_t, the source domain and the target domain sample $\mathbf{X} = \left[\mathbf{X}_s; \mathbf{X}_t^l; \mathbf{X}_t^u \right]$ are projected locally, and the data in the original space is mapped to the D-dimensional manifold space to obtain $\mathbf{Z} = \left[\mathbf{Z}_s; \mathbf{Z}_t^l; \mathbf{Z}_t^u \right]$.

Step 4: The cross-domain density ratio estimation of the source domain samples is calculated as a priori of the transfer weight $\omega_0(z_i^s) \in \mathbb{R}^{n_s \times 1}$. In order to obtain more reliable results, the samples with smaller label errors in the target domains are given 100% weight in the cross-domain transfer to obtain $\hat{\omega}_0 \in \mathbb{R}^{(n_s + n_t^l) \times 1}$, and a more accurate sample weight $\hat{\omega}$ in the source domain is obtained by regularization.

Step 5: Based on $\ell(A, \hat{\omega}, D_s, D_t) = \ell_w(A, \hat{\omega}) - \ell_b(A, \hat{\omega})$, a manifold metric A is found, which can minimize the distribution difference between the source domain and the target domain in manifold space and eliminate the overlap between classes in the domain. Under this metric, the manifold regularization term $\Phi(G)$ is constructed and the manifold structure in the space is fully excavated, so that the cross-domain distance is smaller and the distinction between different classes is more obvious: $Z_s^{new} = Z_s M$ and $Z_t^{new} = Z_t M$.

Step 6: The pseudo-label $\hat{y}_t(z)$ is updated according to the new sample distribution. After each iteration the label propagation range is increased and steps 2 to 6 are repeated until the results converge.

6. Experiments for Validation

6.1. Experiment on Shaft Current Damage Identification of Motor Bearing under Variable Working Conditions

In order to verify that the proposed method has better recognition ability for the typical faults of motor bearings under variable working conditions, the proposed method is compared with several popular transfer learning methods to evaluate the performance of the proposed method.

In this paper, the algorithm validation of a motor bearing shaft current damage simulation test bench is shown in Figure 6. The test bearing used in this experiment is a detachable deep groove ball bearing whose model is 6205 EKA. Three different working conditions (500 N/1800 rpm, 1000 N/1200 rpm, 1500 N/600 rpm) were set in the experiment, and the sampling frequency was 16,384 Hz. The data used for verification have eight health states: normal (NO), outer ring fault (OF), inner ring fault (IF), rolling body fault (BF), and shaft current damage (SCD). There are 100 samples in each state of bearing, each sample contains 2048 data points, and each sample set has 500 samples in total. Three sample sets under different working conditions are obtained by labelling the data sets, and two data sets form a transfer task, so that six transfer tasks of damage identification under variable conditions can be obtained.
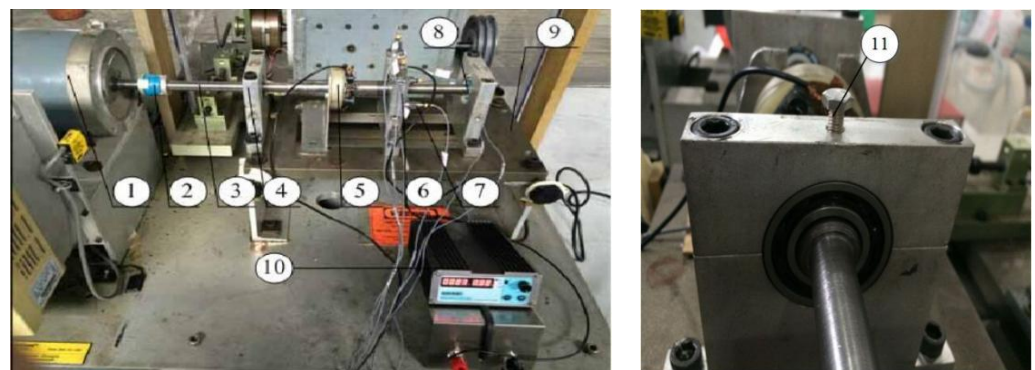

Figure 6. Motor bearing shaft current damage test bench: ① Motor; ② Insulated coupling; ③ Spindle; ④ Bearing housing; ⑤ Carbon brush; ⑥ Test bearing housing; ⑦ Vibration acceleration sensor; ⑧ Bearing; ⑨ Base; ⑩ Power; ⑪ Load bolts.

The data set is shown in Tables 1 and 2.

Table 1. Shaft current damage data set description of bearing.

Data Set	Working Condition	Num of Sample	Class
A	500 N/1800 rpm	500	5
B	1000 N/1200 rpm	500	5
C	1500 N/600 rpm	500	5

Table 2. Type description of bearing data set.

Class	1	2	3	4	5
Type	NO	OF	IF	BF	SCD

Furthermore, part of its signal spectrum is shown in Figure 7. We can find that the spectrum of the same fault type varies greatly under different working conditions, especially the spectrum of shaft current damage, making it difficult to identify faults under variable working conditions.
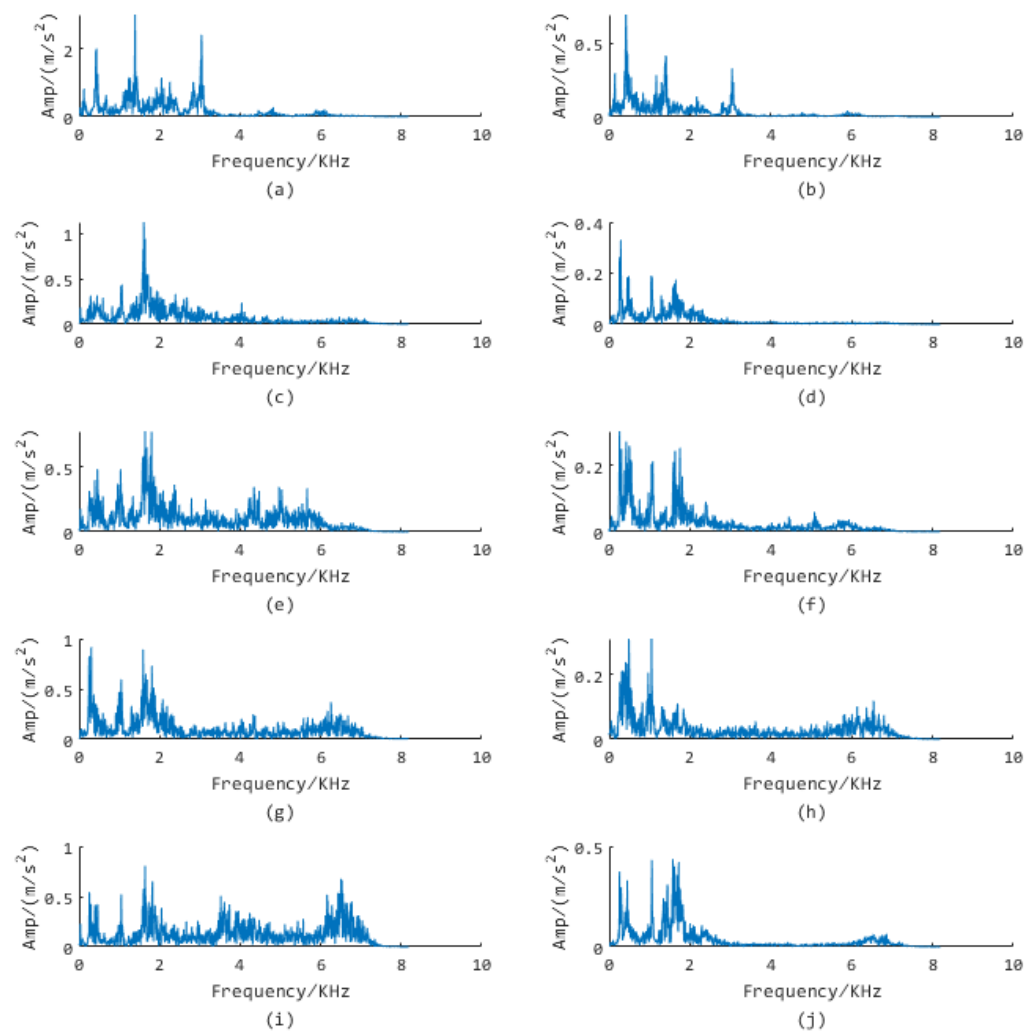
Figure 7. Signal spectrum of motor bearing fault data Set. (**a**) Spectrum of NO under Working Condition A, (**b**) Spectrum of NO under Working Condition C, (**c**) Spectrum of OF under Working Condition A, (**d**) Spectrum of OF under Working Condition C, (**e**) Spectrum of IF under Working Condition A, (**f**) Spectrum of IF under Working Condition C, (**g**) Spectrum of BF under Working Condition A, (**h**) Spectrum of BF under Working Condition C, (**i**) Spectrum of SCD under Working Condition A, (**j**) Spectrum of SCD under Working Condition C.

6.2. Experimental Results

In order to reflect the effectiveness of the proposed method, we present the superiority of the proposed method from two aspects: One is to compare the multiscale sub-band fuzzy entropy with the ordinary time domain and frequency domain features (here we extract 24-dimensional time domain and frequency domain features, such as effective value, standard deviation, kurtosis, and average frequency). The second is to compare the proposed method with other machine learning algorithms and compare two typical traditional algorithms: KNN and SVM. There are two mainstream transfer learning methods, including one unsupervised transfer learning method, MEDA, and one semi-supervised metric transfer learning method, SSMTL [47]. The transfer learning methods all use the KNN algorithm to obtain weak labels for the target domain. The initial subspace dimension is set to 10, and the same hyperparameters involved in the other algorithms are set to the same value. The regularization parameters in the algorithm are determined by optimizing from 0 to 1: $\rho = 1, \lambda = 1, \gamma = 0.01, \sigma = 1$. The subspace dimension $d = 10$, gradient descent step is set to a smaller value: $u_1 = 10^{-5}, u_2 = 10^{-5}$, gradient convergence threshold $\varepsilon = 0.01$, and iteration times $T = 8$, The motor bearing fault identification results of five classification

algorithms under different transfer tasks are obtained as shown in Table 3, which is the result of all tasks.

Table 3. Recognition accuracy of different methods for motor bearing fault under variable conditions.

Feature	Method	Task							
		A→B	A→C	B→A	B→C	C→A	C→B	Average	Std.
Time domain and frequency domain feature	KNN	94.80	66.20	96.40	80.80	74.40	82.80	82.57	10.64
	SVM	95.80	73.20	99.00	81.80	95.40	**96.00**	90.20	9.39
	MEDA	99.00	78.40	99.60	79.80	94.00	82.60	88.90	8.90
	SSMTL *	92.93	78.59	98.18	82.63	86.26	87.47	87.68	**6.43**
	LP-MMT	**99.80**	**83.40**	**100.00**	**94.60**	79.00	**100.00**	**92.80**	8.51
Multiscale sub-band fuzzy entropy	KNN	97.20	84.00	98.00	95.80	87.20	93.80	92.67	5.25
	SVM	99.80	84.60	100.00	94.40	90.40	94.20	93.90	5.34
	MEDA	92.80	92.80	93.80	93.80	98.00	98.00	94.87	2.25
	SSMTL *	99.19	86.26	100.00	98.18	93.94	96.16	95.62	4.64
	MFLP-MMT	**100.00**	**99.80**	**100.00**	**100.00**	**100.00**	**100.00**	**99.97**	**0.07**

* Represents semi-supervised metric transfer learning and the proportion of supervised samples is 1%.

Comparing different features, when the extracted features are the time domain and frequency domain features, the average recognition rate of the migration task of the proposed method is higher than that of other algorithms under variable working conditions. When the extracted features are multiscale sub-band fuzzy entropies, it can be seen that the accuracy of all methods is greatly improved compared with the ordinary time domain and frequency domain features. The MFLP-MMT method in this paper has the highest accuracy in all migration tasks, the average recognition rate is 99.97%, the standard deviation is the lowest, and the stability of the algorithm is the highest. It is proved that the proposed multiscale sub-band fuzzy entropy feature can deeply explore the essential characteristics of bearing vibration signals and can more clearly distinguish between the common bearing faults and current damage faults.

As one of the most basic classification methods, KNN does not change the original distribution of the classification samples and only measures the Euclidean distance between samples, so its average accuracy is always the lowest. As a widely used traditional classification method, the SVM method uses the inner product kernel function instead of the nonlinear mapping to the high-dimensional space, and the idea of maximizing the classification margin is the core of the SVM method. Therefore, SVM performs well in the sample classification of ordinary time-frequency domain features. However, when we extract multiscale features, the transfer learning method still has obvious advantages. As a metric transfer learning method, the SSMTL method measures and eliminates the distribution difference between the source domain and the target domain through the target domain supervision sample, so it can also have higher accuracy. The reason is that the different distributions of the variable condition samples are key problems that need to be solved by transfer learning.

Comparing the results of the transfer tasks A→C and C→A with those of the other tasks, it can be seen that the larger the span of the variable working conditions of the bearing, the worse the fault diagnosis effect. The reason is that the two sets of data with different working conditions have different spatial distributions, and the original space in which they are located covers its manifold characteristics. The larger the span of the working conditions, the greater the difference in the spatial distribution. In particular, for the transfer tasks A→C and C→A with a large span of variable working conditions, the recognition rate is also increased from about 80% to nearly 100%, which can better reflect the superiority of multiscale features. At the same time, the proposed MFLP-MMT method is used to fully mine the internal relationship of the data in the manifold space. Considering the inter-domain data offset, the source domain training samples are weighted. Further adjusting the distance between the classes in the domain can effectively avoid the

overlap in the domain and adjust the spatial distribution of the data through iteration. The results show that the proposed method can effectively and accurately identify the multiple fault types of motor bearings under variable working conditions, and has a good performance, which is about 5% higher than the widely used unsupervised transfer learning method MEDA. The SSMTL method has high accuracy in tasks with a small migration span, but it is not effective when the working condition span is too large, so this is also its deficiency. As an unsupervised transfer learning algorithm, this method does not need to supervise the samples of the target working conditions, and all the damage identification tasks under variable working conditions are close to 100%, which reflects the superior performance of this method in the damage identification of bearing shaft current under variable working conditions.

6.3. Experimental Analysis

6.3.1. Comparison between Manifold Space and Linear European Space

In order to reflect the importance of manifold mapping in this method, we compare the manifold mapping method LPP used in this method with the traditional linear dimensionality reduction method principal components analysis(PCA). By changing the dimension of the target projection space, the extracted multiscale sub-band fuzzy entropy features are implicitly projected into the target space to show the necessity of low-dimensional manifolds, and the experimental results are shown in Figure 8.
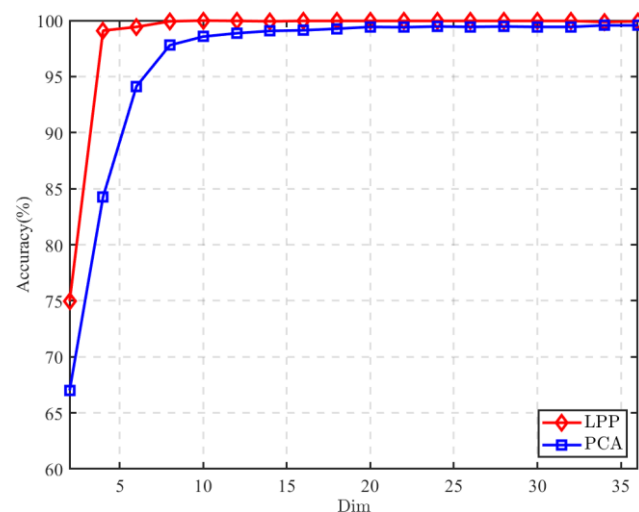

Figure 8. Influence of different dimensionality reduction methods on fault recognition rate.

From the figure, it can be seen that after the dimension reduction by LPP manifold, the fault recognition rate in all dimensions is higher than that of the traditional linear dimension reduction method PCA, and the advantage is more obvious when the spatial dimension is lower. In theory, the common fault of the same part of the motor bearing is very similar to the bearing current damage fault in the original space. As the spatial dimension changes, it can be found that the low-dimensional manifold space contains more sample information than the low-dimensional Euclidean space. As the spatial dimension changes, it can be found that the low-dimensional manifold space contains more sample information than the low-dimensional Euclidean space. When the manifold space dimension is about eight, the fault recognition accuracy is basically the highest. Using the linear dimension reduction PCA method, the fault recognition accuracy is positively correlated with the spatial dimension, indicating that the low-dimensional original space will lose sample information, indicating that the low-dimensional manifold mapping can indeed mine the hidden manifold structure in the original space. At the same time, the space with lower dimensions can reduce the complexity of the subsequent algorithm and reduce the computational cost.

6.3.2. Convergence Analysis of Algorithm

In order to study the convergence of the proposed method in the iterative process, we analyze the convergence of the algorithm for all variable working condition transfer tasks in the motor bearing fault data set and compare the two characteristics of the ordinary time domain frequency characteristics and multiscale sub-band fuzzy entropy. The results are shown in Figure 9.
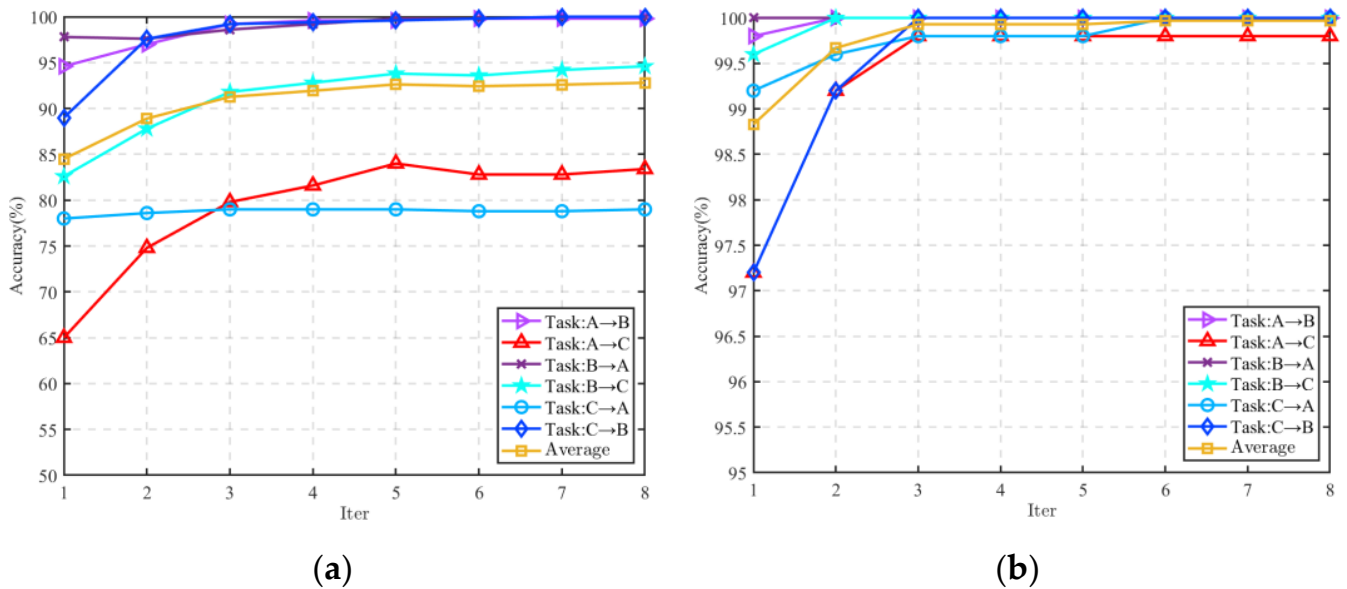

Figure 9. Convergence process of this algorithm: (**a**) time and frequency domain characteristics; (**b**) multiscale sub-band fuzzy entropy.

It can be seen from the figure that when the proposed features are the time domain and the frequency domain, some transfer tasks with small operating ranges can achieve more than 90% or even 100% accuracy through iteration. For the task A→C and task C→A with the large cross-domain transfer, because the ordinary time-domain and frequency-domain features cannot reflect the essential information of the vibration signal, when the operating conditions of the bearing are too different, the external information will be reflected. The characteristics of the vibration fault signal are then submerged, which makes the recognition rate of the fault classification method decrease significantly.

We analyze the recognition results of task A→C and task C→A through the confusion matrix. As shown in Figure 10, we can see that the shaft current damage fault recognition rate of the motor bearing is the lowest in both transfer tasks. This is due to the damage process of the bearing current. It is a gradual evolution process. In the early stage of the damage, it is difficult to identify the shaft current damage fault of the bearing because the fault signal is too weak and the working condition changes too much. With the aggravation of the shaft current damage fault, the shaft current will form an electrical damage fault on the outer ring, ball, and inner ring of the motor bearing successively. Because the fault characteristics are not obvious, it is easy to confuse this kind of electrical damage fault with other common faults, which affects the maintenance and the repair of the equipment in the subsequent engineering practice.
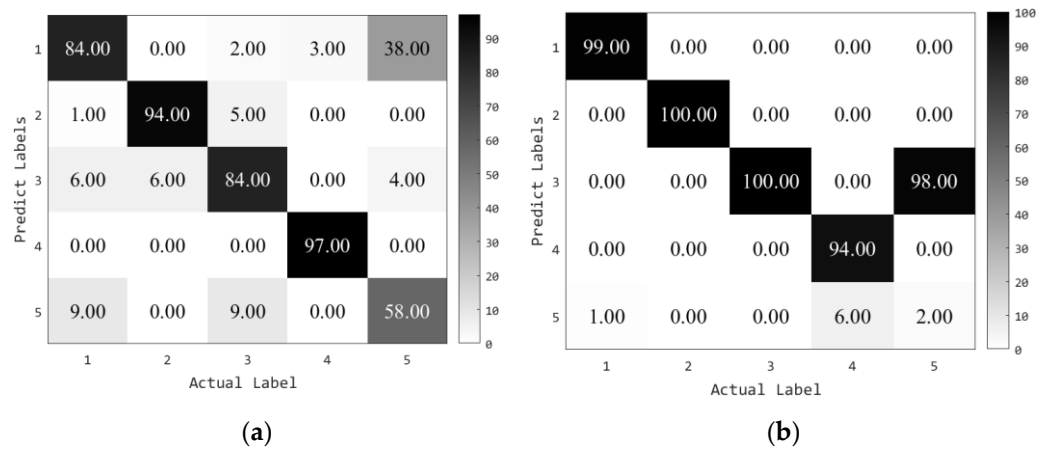
Figure 10. Identification results after extracting time domain and frequency domain features: (**a**) task A→C; (**b**) task C→A.

When the proposed feature is multiscale sub-band fuzzy entropy, the accuracy of the small part of the transfer task in the source domain and the target domain can reach more than 98% in the first iteration. Only one iteration can obtain a better recognition performance, and the recognition rate can be improved to about 100% after 2–3 iterations. Even if the sample label error after the first iteration is large, it can quickly converge to 100% during the iteration, as shown in Figure 11 which is the transfer task C→B iterative convergence process. Furthermore, for all the transfer tasks in the iterative process, there is no significant decline in the identification accuracy, indicating that the proposed method has a good convergence.
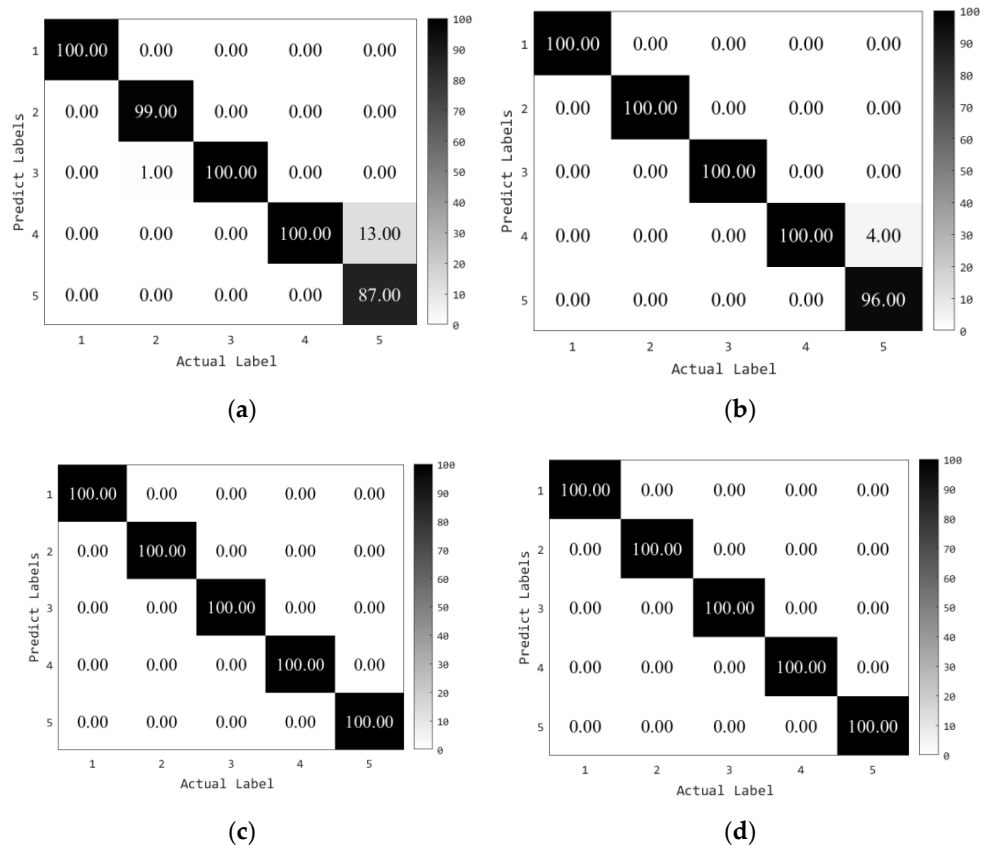

Figure 11. Convergence process of transfer task C→B: (**a**) first iteration; (**b**) second iteration; (**c**) third iteration; (**d**) last iteration.

6.3.3. Visual Analysis

In order to observe the distribution of the data more intuitively, t-distribution stochastic neighbor embedding (t-SNE) algorithm is used to visualize the distribution of the fault types of tasks A→C and C→B and present the effect of the manifold measurement in the form of a two-dimensional scatter graph. As shown in Figure 12, "□" represents the source domain sample and "×" represents the target domain sample.
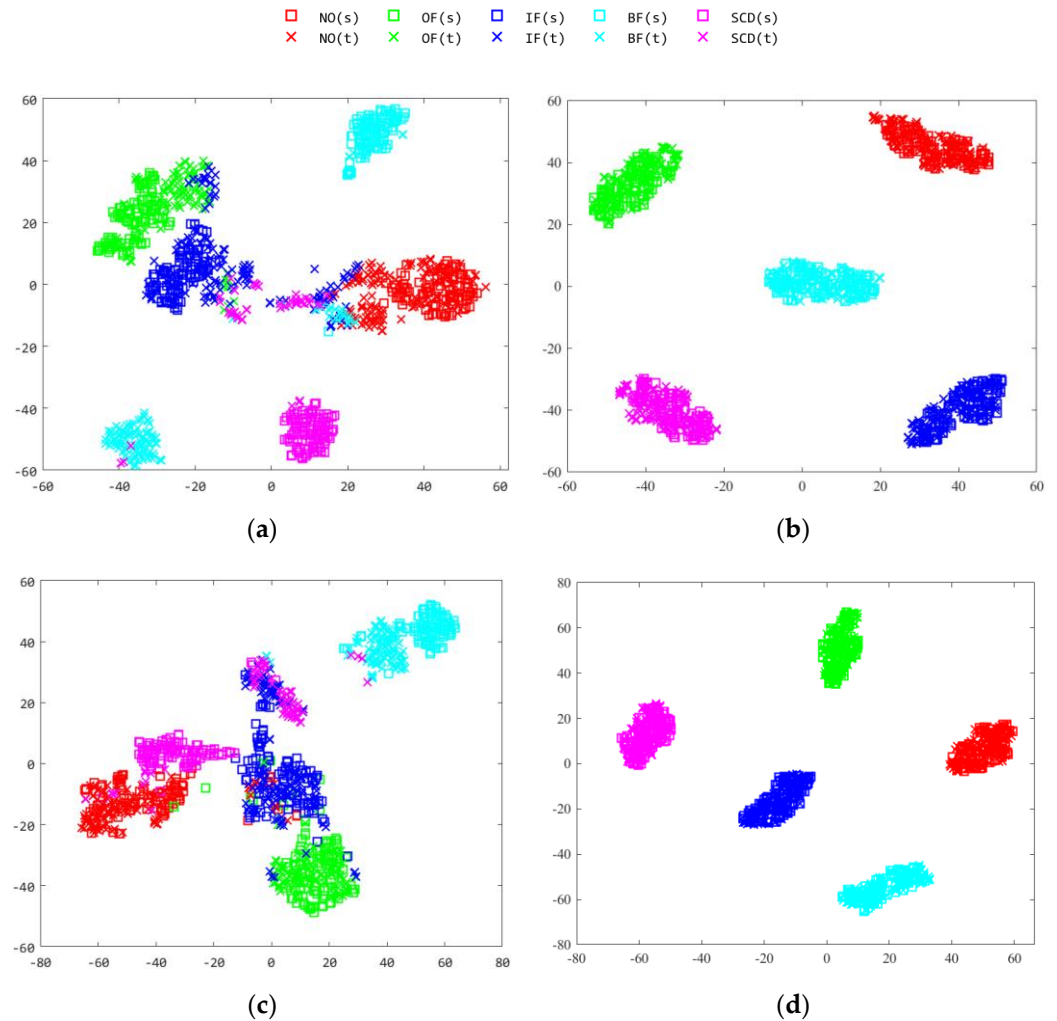

Figure 12. Two-dimensional scatter plot of bearing fault sample distribution. (**a**) Original distribution of samples (Task: A→C); (**b**) MFLP-MMT (Task: A→C); (**c**) original distribution of samples (Task: C→B); (**d**) MFLP-MMT (Task: C→B).

It can be seen from the figure that, in the original space, the distribution of the fault type data under variable working conditions is quite different, and the overlap of the different types of motor bearing faults is serious. If the traditional machine learning algorithm is directly used for fault classification, it is easy to confuse the different fault types, especially for the motor bearing shaft. The recognition effect of the current damage is very poor. The proposed method can effectively reduce the cross-domain offset, while the same class of data narrows the distribution distance, achieving the effect of clustering and the distribution of different classes of samples of alienation, eliminating the overlap between classes within the domain, and achieving accurate identification of motor bearing shaft current damage.

7. Conclusions

In this paper, a method based on multiscale feature label propagation and manifold metric transfer is proposed to identify the shaft current damage of motor bearings under variable working conditions. By extracting multiscale sub-band fuzzy entropy, the essential characteristics of the weak damage signal of the motor bearings are deeply excavated. At the same time, it makes up for the shortcomings of the existing transfer learning methods in the fault diagnosis of bearing under variable working conditions, such as only the cross-domain distribution alignment, ignoring the distribution relationship between classes in the domain, and ignoring the label information in the domain. It solves the problems that the existing metric learning methods only focus on the metric of the original space and do not take into account the manifold relationship between the samples, and breaks the limitations of the need to supervise the samples in the process of the metric. By finding the manifold metric with the smallest cross-domain distance and obvious alienation of heterogeneous samples in the manifold space, the intra-domain overlap phenomenon of typical damage types of motor bearings is solved. Based on the graph label propagation method, the limited label information is gradually diffused in the manifold space to realize the cross-domain propagation of labels, and the problem of the low recognition accuracy of motor bearing shaft current damage under variable working conditions is solved. The experimental results show that the proposed method has a good recognition ability for typical faults of motor bearings and a strong convergence. The average accuracy of the diagnosis model is close to 100%. In industrial practice, different maintenance measures are taken, respectively, through the accurate diagnosis of the common faults and shaft current damage faults of the motor bearings. Moreover, the high accuracy of the fault diagnosis under variable working conditions greatly reduces the cost of the training diagnosis model. Only a labeled sample under one working condition can be collected, and the samples under other different working conditions can be efficiently identified.

In future work, we will continue to explore the recognition ability of MFLP-MMT in motor bearing current damage identification such as bearing current damage diagnosis for different types of motors.

Author Contributions: Conceptualization, G.W. and Z.Z.; methodology, S.Z.; software, S.Z.; validation, G.W., S.Z. and Z.Z.; formal analysis, D.Z.; investigation, G.W.; resources, G.W.; data curation, D.Z.; writing—original draft preparation, S.Z.; writing—review and editing, G.W.; visualization, S.Z.; supervision, G.W.; project administration, G.W.; funding acquisition, G.W. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data that support the findings of this study are available upon reasonable request from the authors.

Conflicts of Interest: This manuscript has not been published elsewhere, and it has not been submitted simultaneously for publication elsewhere. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

1. Yan, G.; Zhi, Y.; Chen, X.; Zhong, W. Research on the principle and suppression technology of electric corrosion of traction motor bearings for metro vehicles. *Electr. Drive Locomot.* **2019**, *4*, 102–106.
2. Erdman, J.M.; Kerkman, R.J.; Schlegel, D.W.; Skibinski, G.L. Effect of PWM inverters on AC motor bearing currents and shaftvoltages. *IEEE Trans. Ind. Appl.* **1996**, *32*, 250–259. [CrossRef]
3. Chen, S.; Lipo, T.A.; Fitzgerald, D. Modeling of motor bearing currents in PWM inverter drives. *IEEE Trans. Ind. Appl.* **1996**, *32*, 1365–1370. [CrossRef]
4. Chen, G.; Chen, G.; Xu, M. Wind turbine shaft voltage shaft current impact on the bearing and improvement measures. *Shenhua Technol.* **2014**, *12*, 60–62.
5. Xu, W.; Li, D.; Yang, K. Shaft current generation and bearing protection of hydro-generator. *China Equip. Eng.* **2019**, *5*, 51–52.

6. Adabi, M.E.; Abolfazl, V. A survey of shaft voltage reduction strategies for induction generators in wind energy applications. *Renew. Energy* **2013**, *10*, 177–187. [CrossRef]

7. Prashad, H. Appearance of craters on track surface of rolling element bearings by spark erosion. *Tribol. Int.* **2001**, *34*, 39–47. [CrossRef]

8. Liu, W. The prevalent motor bearing premature failures due to the high frequency electric current passage. *Eng. Fail. Anal.* **2014**, *45*, 118–127. [CrossRef]

9. Picot, A.; Obeid, Z.; Régnier, J.; Poignant, S.; Darnis, O.; Maussion, P. Statistic-based spectral indicator for bearing fault detection in permanent-magnet synchronous machines using the stator current. *Mech. Syst. Signal Process.* **2014**, *6*, 424–441. [CrossRef]

10. Liu, R.; Meng, Y.; Ren, X.; Wang, Q. Analysis of Shaft Current Problem of Direct Drive Permanent Magnet Synchronous Wind Generator. *J. Electr. Mach. Control* **2019**, *23*, 43–49.

11. Ren, X.; Liu, R.; Wang, Q. Investigation of bearing model and calculation of equivalent resistance in shaft current problems. *J. Electr. Eng.* **2019**, *2*, 1–6.

12. Kempski, A.; Robert, S.; Jacek, B. Statistical model of electrostatic discharge hazard in bearings of induction motor fed by inverter. *J. Electrost.* **2005**, *63*, 475–480.

13. Houssin, E.L.; Bouchikhi, E.L.; Choqueuse, V. Induction machine faults detection using stator current parametric spectral estimation. *Mech. Syst. Signal Process.* **2015**, *52–53*, 447–464.

14. Wang, G.; He, Y.; Peng, Y.; Li, H. Bearing Fault Identification Method based on Collaborative Filtering Recommendation Technology. *Shock Vib.* **2019**, *2019*, 7378526. [CrossRef]

15. Wang, G.; Du, M.; Huang, L.; Li, L. Prediction of bearing damage in wind turbines based on the quadratic root mean square of sub-band manifold. *J. Mech. Eng. Sci.* **2017**, *10*, 3213–3223.

16. Chen, B.; Qing, T.; Cao, X.; He, W.; Zeng, N. Application of lifting wavelet fractal strategy in feature extraction of electrical erosion fault of motor bearing. *Vib. Shock* **2022**, *41*, 223–230, 247.

17. Muetze, A.; Tamminen, J.; Ahola, J. Influence of Motor Operating Parameters on Discharge Bearing Current Activity. *IEEE Trans. Ind. Appl.* **2011**, *47*, 1767–1777. [CrossRef]

18. Schuster, M.; Springer, J.; Binder, A. Comparison of a 1.1 kW-induction machine and a 1.5 kW -PMSM regarding common-mode bearing currents. In Proceedings of the 2018 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM), Amalfi, Italy, 20–22 June 2018; pp. 1–6.

19. Khan, K.; Gyllensten, F. Experimental investigation of bearing currents in low voltage motors. In Proceedings of the 2018 XIII International Conference on Electrical Machines (ICEM), Alexandroupoli, Greece, 3–6 September 2018; pp. 218–224.

20. Magdun, O.; Gemeinder, Y.; Binder, A. Investigation of influence of bearing load and bearing temperature on EDM bearingcurrents. In Proceedings of the 2010 IEEE Energy Conversion Congress and Exposition, Atlanta, GA, USA, 12–16 September 2010; pp. 2733–2738.

21. Suzumura, J. Prevention of electrical pitting on rolling bearings by electrically conductive grease. *Q. Rep. RTRI* **2016**, *57*, 42–47. [CrossRef]

22. Gonda, A.; Capan, R.; Bechev, D.; Sauer, B. The Influence of Lubricant Conductivity on Bearing Currents in the Case of Rolling Bearing Greases. *Lubricants* **2019**, *7*, 108. [CrossRef]

23. Inoue, E.; Mitsuoka, M.; Rabbani, M. Investigation of nonlinear vibration characteristics of agricultural rubber crawler vehicles. *Ama-Agric. Mech. Asia Afr. Lat. Am.* **2011**, *42*, 89.

24. Zhuang, F.; Luo, P.; He, Q. Advances in transfer learning. *J. Softw.* **2015**, *26*, 26–39.

25. Long, J.; Chen, Y.; Yang, Z.; Huang, Y.; Li, C. A novel self-training semi-supervised deep learning approach for machinery fault diagnosis. *Int. J. Prod. Res.* **2022**, 1–14. [CrossRef]

26. Li, X.; Li, Y.; Yan, K.; Shao, H.; Lin, J. Intelligent fault diagnosis of bevel gearboxes using semi-supervised probability support matrix machine and infrared imaging. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108921.

27. Shen, B.; Chen, B.; Zhao, C. Review of deep learning in mechanical equipment fault prediction and health management. *Mach. Tools Hydraul.* **2021**, *49*, 162–171.

28. Wang, H.; Geng, L.; Ni, T. Knowledge-embedded transfer twin support vector machine. *Control Decis.* **2019**, *34*, 519–526.

29. Zhang, H.; Han, X.; Shi, K.; Ren, H.; Xu, L.; Qian, H.; Li, C.; Li, F.; Huang, H. Research on online fault diagnosis system of hoist spindle based on deep transfer learning. *Coal Eng.* **2022**, *54*, 61–66.

30. Chen, R.; Tang, L.; Hu, X.; Yang, L.; Zhao, L. Rolling bearing fault diagnosis method based on deep attention transfer learning at different speeds. *Vib. Shock* **2022**, *41*, 95–101, 195.

31. Xu, H.; Zhang, W.; Song, D.; Chen, B. Research on SSDAE Rolling Bearing Fault Diagnosis Method Based on Deep Transfer. *Noise Vib. Control* **2021**, *41*, 112–118.

32. Lan, Y.; Hu, C.; Jin, J.; Wang, Y. Research on cross-domain bearing fault classification method based on transfer component analysis. *Electromech. Eng.* **2021**, *38*, 521–527.

33. Wang, J.; Feng, W.; Chen, Y. Visual domain adaptation with manifold embedded distribution alignment. In Proceedings of the 2018 ACM Multimedia Conference, Seoul, Republic of Korea, 22–26 October 2018; pp. 402–410.

34. Long, M.; Wang, J.; Ding, G.; Sun, J.; Yu, P.S. Transfer feature learning with joint distribution adaptation. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2200–2207.

35. Zhang, X.; Yu, D.; Liu, S. Research on small sample bearing fault diagnosis method based on transfer learning. *J. Xi'an Jiaotong Univ.* **2021**, *55*, 30–37.
36. Zhao, K.; Jiang, H.; Wu, Z. A novel transfer learning fault diagnosis method based on Manifold Embedded Distribution Alignment with a little labeled data. *J. Intell. Manuf.* **2022**, *33*, 151–165. [CrossRef]
37. Zhi, H.; Yu, H.; Li, S. A deep metric learning based video classification method. *J. Electron. Inf. Technol.* **2018**, *40*, 27–34.
38. Xu, Y.; Ma, J.; Chen, L.; Shen, C.; Li, Q.; Kong, L. Bearing Fault Diagnosis Based on Similarity Measure Transfer Learning. *Vib. Shock* **2022**, *41*, 217–223.
39. Zhao, X.; Wang, Y.; Zhang, Y.; Wu, J.; Wang, L. A compound fault identification method of gearbox based on Triplet loss. *Vib. Shock* **2021**, *40*, 46–54.
40. Huang, N.; Yang, X.; Cai, G. Diagnosis of fan main bearing fault depth confrontation using unbalanced small sample data. *Chin. J. Electr. Eng.* **2020**, *40*, 563–574.
41. Jin, R.; Wang, S.; Zhou, Y. Regularized distance metric learning: Theory and algorithm. In *Advances in Neural Information Processing Systems*; Curran: Red Hook, NY, USA, 2009; pp. 862–870.
42. Cao, B.; Ni, X.; Sun, J.T.; Wang, G.; Yang, Q. Distance metric learning under covariate shift. In Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011; pp. 1204–1210.
43. Seyedi, S.A.; Lotfi, A.; Moradi, P. Dynamic graph-based label propagation for density peaks clustering. *Expert Syst. Appl.* **2019**, *115*, 314–328. [CrossRef]
44. He, X.; Niyogi, P. Locality preserving projections. In *Advances in Neural Information Processing Systems*; Curran: Red Hook, NY, USA, 2003.
45. McClure, K.S.; Gopaluni, R.B.; Chmelyk, T. Nonlinear process monitoring using supervised locally linear embedding projection. *Ind. Eng. Chem. Res.* **2014**, *53*, 5205–5216.
46. Sugiyama, M.; Nakajima, S.; Kashima, H.; Buenau, P.V.; Kawanabe, M. Direct importance estimation with model selection and its application to covariate shift adaptation. In *Advances in Neural Information Processing Systems*; Curran: Red Hook, NY, USA, 2008; pp. 1433–1440.
47. Sanodiya, R.K.; Mathew, J. A framework for semi-supervised metric transfer learning on manifolds. *Knowl.-Based Syst.* **2019**, *176*, 1–14. [CrossRef]