IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 13, 2021, accepted March 1, 2021, date of publication March 4, 2021, date of current version July 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3063743

Cross-Domain Intelligent Fault Diagnosis Method of Rotating Machinery Using Multi-Scale Transfer Fuzzy Entropy

ZHENG DANGDANG[1,2], BING HAN[1], GENG LIU[1], YONGBO LI[1], (Member, IEEE), AND HUANGCHAO YU[3]

[1]Shaanxi Engineering Laboratory for Transmissions and Controls, Northwestern Polytechnical University, Xi'an 710072, China
[2]First Aircraft Institute, Aviation Industries of China (AVIC), Xi'an 710089, China
[3]College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China

Corresponding author: Huangchao Yu (huangcha@ualberta.ca)

This work was supported in part by the National Natural Science Foundation of China under Grant 51805414 and Grant 51905537.

ABSTRACT A novel rotating machine fault diagnosis method based on multi-scale transfer fuzzy entropy (MTFE) and support vector machine (SVM) is proposed in this paper. Compared with traditional machine learning methods, our proposed method can identify various fault types of rotating machinery with different data distribution by learning the transfer knowledge from different distribution source domains. First, multi-scale fuzzy entropy (MFE) features of all samples are extracted as input. Second, build a transfer learning model to find a projection matrix to map the MFE features of the source and target domains into a common subspace, called MTFE features. In this process, the distribution structure of training data is maintained and the distribution difference between training data and test data is reduced. Finally, the SVM classifier identifies the fault type of the test data. Two cases including gearbox and rolling bearing are used for validation. Experimental results demonstrate that our proposed MTFE method performs best in recognizing various fault types comparing with other six methods.

INDEX TERMS Cross-domain fault diagnosis, rotating machinery, transfer learning, fuzzy entropy.

NOMENCLATURE

ApEn	approximate entropy
CWRU	Case Western Reserve University
FE	fuzzy entropy
KNN	k-nearest neighbour
K-S entropy	Kolmogorov-Sinai entropy
LR	logistic regression
MFE	multi-scale fuzzy entropy
MFPT	Machinery Failure Prevention Technology
MMD	maximum mean difference
MMDE	maximum mean difference embedding
MTFE	multi-scale transfer fuzzy entropy
PCA	principal component analysis
PE	permutation entropy
PHM	prognostic health management
PHM09	Prediction and Health Management Society for the 2009 data challenge
SDE	symbolic dynamic entropy
SE	sample entropy
SSTCA	semi-supervised transfer component analysis
SVM	support vector machine
TCA	transfer component analysis
TL	transfer learning
TLPP	transfer locality preserving projection

The associate editor coordinating the review of this manuscript and approving it for publication was Filbert Juwono.

I. INTRODUCTION

The prognostic health management (PHM) ensures that the machinery can run normally, which has great significance to reduce economic losses and guarantee operation safety [1]. Rotating machinery is widely used in industrial production. Therefore, plenty of intelligent fault diagnosis algorithms have been developed to identify fault types of important components such as rolling bearings and gearboxes.

The artificial intelligent bases fault diagnosis is implemented by the steps of data collection, feature extraction, and health state recognition according to Ref [1]. By machine learning theories, the diagnosis models are able to automatically recognize the health conditions of machines. With the rapid development of machine learning over the recent

years, the advent of deep learning brings positive effects on the enhanced benefits. Intelligent fault diagnosis approaches mainly contain the following categories: expert system-based approaches [2], SVM-based approaches [3], and deep learning-based approaches [4]. However, it should be concerned that the successes of intelligent fault diagnosis models are subject to enough labeled samples. Such assumption is unpractical in engineering scenarios. To bridge the gap, transfer learning theories are promising to construct diagnosis models, in which the diagnosis knowledge can be transferred across multiple diagnosis tasks.

Among traditional intelligent fault diagnosis models, the quality of the features directly affects the classification results. There are many feature extracting algorithms, such as time-frequency domain features, complexity-based features, features extracted based on artificial intelligence algorithms, and so on [5]–[7]. Due to the complex working environment, the nonlinear stiffness of the rolling body and other factors, the vibration signal of rotating machinery is usually nonlinear and non-stationary. In the diagnostic process of rotating machinery, the traditional time and frequency domain feature extraction techniques for stationary linear signals are always difficult to extract effective features [8]. At this time, the nonlinear signal quantitative description method, which represents the complexity of the signal [9], gradually enters the horizon of researchers. Entropy can measure the disorder degree of time series, so it can represent the abrupt behavior of system dynamics when rotating machinery fails [10]–[12]. Kolmogorov proposed K-S entropy in 1958, which was used to depict the complexity of a system. Developed on the K-S entropy [13], Pincus proposed approximate entropy (ApEn) [14], which measures the complexity of a signal by taking into account the rate at which new information is generated in a time series. Then, Costa proposed sample entropy (SE) to resolve the issue of self-matching in ApEn [15]. Chen introduces the fuzzy membership function into the sample entropy and proposes a more accurate fuzzy entropy (FE) [16]. In addition, some researchers symbolized the time series to observe its permutation pattern, and typical studies include permutation entropy (PE) and symbolic dynamic entropy (SDE) [17], [18].

Combined with the multi-scale idea proposed by Costa *et al.* [19], Zheng proposed the multi-scale fuzzy entropy (MFE), which can extract the information of more scales in vibration signals [20], [21]. The application of existing entropy-based methods only provided the feature extraction algorithm, which requires a large quantity of labelled data to train the intelligent model for classification. The above-mentioned entropy-based method is appropriate to process the training and testing data following the same distribution. This means we need to obtain data of the same distribution prior to diagnosis. However, there are lots of difficulties in obtaining such data in practical tasks:

(a) The device is not allowed to run for long periods of time in a faulty state, and it is difficult to obtain large amounts of data.

(b) The equipment may fail with only one type of failure, and all types of fault data cannot be accumulated.

(c) It is not possible to implement simulations on every device to be analyzed.

Therefore, in a real-world scenario, it is difficult to obtain a sufficient amount of labelled data to train the model. Model can only be trained with data that does not meet the same distribution assumptions. This is the cross-domain fault diagnosis problem that this paper focuses on.

To solve the problem, many researchers are trying to find further mathematical transformation of features to reduce the distribution difference between training data and test data, and then input them into classifiers, which is the idea of transfer learning (TL) based on features.

Pan and Yang *et al.* proposed maximum mean difference embedding (MMDE) algorithm to find a kernel function that minimizes the maximum mean difference (MMD) between the training and the test data in a higher dimensional Hilbert space [22], [23]. Then transfer component analysis (TCA) method is evolved on the basis of MMDE, but the difference is that TCA consider the correlation of labels and learn a dimensional reduction matrix to minimize the MMD of the projected data [24]. Zheng et al proposed transfer locality preserving projection (TLPP) algorithm in 2017, which retains the distribution structure of test data and training data while minimizing MMD and achieves high accuracy rate in cross-domain rotating machinery fault diagnosis experiments [25]. These transfer learning methods make it possible to accomplish identify the fault type by different distributed data.

In this paper, the idea of transfer learning is combined with the theory of entropy value, and a method of transfer entropy is proposed, which can extract the common characteristics of the entropy value of test data and training data. First, the MFE features of vibration signal are extracted. Second, MFE features are used as input to construct a model to learn a projection matrix, so that the features after projection keep the spatial adjacent structure, and the MMD value is the minimum. Third, after the model parameters are optimized using bayesian optimization method, output the features after projection, which is called multi-scale transfer fuzzy entropy (MTFE). Finally, the MTFE feature of the training data was used to train the support vector machine (SVM) classifier, judge the fault type of the test data and calculate the recognition accuracy. The contributions of this paper are mainly as follows:

(1) In this paper, a new data-driven fault diagnosis method is proposed, which extract the MTFE features and then use SVM classifier to accomplish fault identification. By extracting features with transfer capability, the diagnostic model can be trained with data that has different distribution, so as to solve the problem of insufficient data.

(2) The feature dimension of MFE is also added to the parameter optimization process, which can help improve the diagnostic accuracy.

(3) The fault diagnosis experiment of gearbox and rolling bearing shows that compared with other traditional machine learning methods, MTFE-SVM method can obtain higher diagnostic accuracy. It means that the method proposed in this paper has a good practical application prospect.

The organization of the rest of this paper is as follows. The MFE procedure and the discussion of its transfer ability are introduced in Section II. The detailed steps of MTFE are provided in Section III. Section IV contains two sets of diagnostic tests for gearbox and rolling bearing and analyses the results. Finally, Section V draws the conclusion of this paper.

II. THEORY BASIS

The problem definition and theory basis for MTFE algorithm are provided in this chapter.

A. PROBLEM DEFINITIONS

First, the problem of fault diagnosis in the actual scenario is defined. Following [26], define the source domain as $\mathcal{D}_s = \{\mathcal{X}_s, P_s(X_s)\}$ and the target domain as $\mathcal{D}_t = \{\mathcal{X}_t, P_t(X_t)\}$, then give source tasks and target tasks as follows: $\mathcal{T}_s = \{\mathcal{Y}_s, f_s\}$, $\mathcal{T}_t = \{\mathcal{Y}_t, f_t\}$. f represents the mapping between the label and the data, f_s is known and f_t need to be learned by using source domain.

In order to simulate the actual fault diagnosis scenario as much as possible, in this paper, the following two points need to be noted. First, dataset $X_s = \{\boldsymbol{x}_{si}\}_{i=1}^{n}$ and $X_t = \{\boldsymbol{x}_{ti}\}_{i=1}^{m}$ are collected from different laboratory, which mean $P_s(X_s) \neq P_t(X_t)$. This is because it is expensive and difficult to collect data with the same distribution as the target domain data in the actual situation, so other different but similar data are needed to make up for the lack of data. Second, the data for the training model consists of the source domain and part of the normal sample of the target domain, and the test data consists of the target domain. This is because in the actual situation, the normal state sample of the equipment to be diagnosed can be obtained, and this part of data is added to the training model, which is also helpful to identify the fault type. Third, the type of failure in the test data should be equal or less than the training data.

B. TRANSFER ABILITY OF MFE FEATURES

In the transfer learning theory, the source domain and the target domain need to be similar, if the difference is too large, it is difficult to learn useful knowledge. MFE features have such a transfer capability, as shown in Figure 1. First, the MFE feature of a sample with scale τ is an τ-dimensional vector. Then, the angle of vectors representing the MFE features of source domain samples collected from different fault types in space is represented by α, as shown in Figure 1(a). Similarly, the sample Angle of the target domain is denoted by β. In the experiment, although the target domain and the source domain are collected from different data sets, the two angles α and β are almost equal. Figure 1(b) shows the MFE feature visualization of CWRU data and MFPT data, including three types of samples: normal (N), outer ring failure (OR) and inner ring failure (IR). Obviously, the spatial structure of the features of the two data sets is very similar. This means that it is feasible to design a dimensionality reduction method that can maintain the spatial structure of the source domain and then apply it to the dimensionality reduction of the target domain.

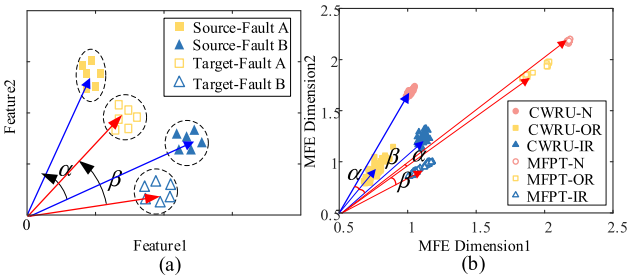

FIGURE 1. (a)Schematic diagram of similarity of spatial structure (b) Structural similarity of real data.

The main purpose of discussing feature similarity of MFE is to accomplish cross-domain subspace learning under the absence of labels in the target domain. In the actual diagnostic scenario, only normal sample of the target domain can be used to train the model, as described in section II.A. Therefore, there are not different types of fault samples to limit the learning process and reduce the distribution differences between domains. Based on this feature correlation, we can retain the possible distribution structure of fault data across domains in training MTFE model process. This allows us to apply the projection model derived from the training data to the test data. Finally, the test samples that do not appear in the training data can be mapped to the correct categories in the MTFE subspace.

III. PROPOSED MTFE METHOD

The detailed process and steps of the MTFE method are illustrated in this chapter.

A. EXTRACT MULTI-SCALE FUZZY ENTROPY

The essential process of fault diagnosis is to extract the features of vibration signals. For normal entropy method, the complexity is only described on a single scale, and a lot of important information is lost. Therefore, it is necessary to extract hidden feature information from different scales. Using the idea of multi-scale entropy for reference, Zheng *et al.* proposed the MFE based on the fuzzy entropy. The calculation steps are described as follows:

(1) Given a time series $\{x_i\} = \{x_1, x_2, \cdots, x_N\}$ and three parameters, embedding dimension m, similar tolerance r and scale factor τ. First, the original sequence is segmented to obtain the coarse-grained time series $\left\{y_j^{\tau}\right\}$ as in (1).

$$y_j^{\tau} = \frac{1}{\tau} \sum_{i=j}^{j+\tau-1} x_j 1 \leq j \leq n - \tau + 1 \tag{1}$$

(2) According to the [8], the fuzzy entropy of each $\left\{y_j^{\tau}\right\}$ is calculated as in (2). Then FE of all the scale factors are

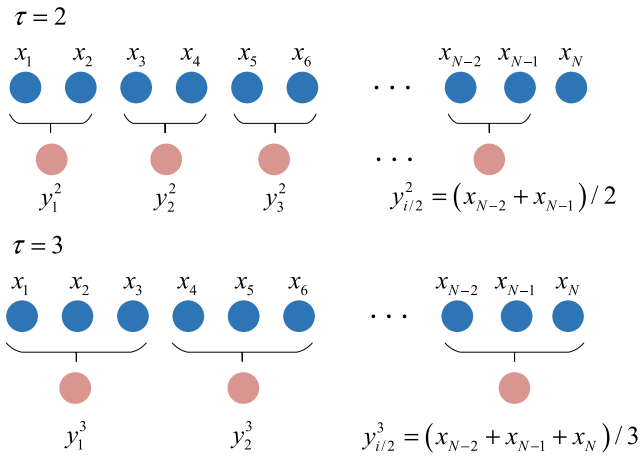

FIGURE 2. Schematic diagram of coarse granulation time series.

combined into MFE.

$$FE(y_j^{\tau}, \boldsymbol{m}, r, N) = -\ln \frac{\Phi^{\boldsymbol{m}+1}(r)}{\Phi^{\boldsymbol{m}}(r)} \tag{2}$$

$$MFE(x, \boldsymbol{m}, r, \tau) = FE(y_j^{\tau}, \boldsymbol{m}, r, N) \tag{3}$$

B. CONSTRUCT TRANSFER LEARNING MODEL

Suppose the MFE features of given source domain and target domain are $\{\boldsymbol{x}_i\}_{i=1}^{n+m}$. Note that n and m represent the number of samples from the source domain and target domain, respectively. A transfer learning model should be construct, which use training data to find a nonlinear mapping matrix, and then project the MFE features of source domain and target domain into the new MTFE subspace. In this paper, the objective function of transfer learning can be described as in (4):

$$L = \mathcal{L}_{LPP} + \lambda \mathcal{L}_{MMD} + \mu \mathcal{L}_R \tag{4}$$

where $\lambda > 0$ is MMD tradeoff parameter, $\mu > 0$ is regularization parameter. $\mathcal{L}_{LPP}$ is used to remain the adjacent structure of projected data. The function of $\mathcal{L}_{MMD}$ is to reduce the distance of normal sample from the target domain and the source domain (This is because the limited training data contains only all the samples of source domain and the normal sample of target domain. If more target domain sample is available, it can also restrict the distance between them, and samples have same labels in the source domain). $\mathcal{L}_R$ is the regularization item, which can prevent model overfitting.

(1) Description of $\mathcal{L}_{LPP}$ term

$\mathcal{L}_{LPP}$ is constructed based on locality preserving projections (LPP) [17] algorithm. The purpose of LPP is to reduce the dimension of the data while retaining the local structure of the data after dimension reduction. In short, adjacent samples in the original sample space still maintain this adjacency relationship after the projection. By learning a transformation matrix $A = [a_1, a_2, \cdots, a_l]_{d\times l}$, the sample is mapped to a new l-dimensional space z (represented by $z(\boldsymbol{x}_i) = A^T \boldsymbol{x}_i \in \mathbb{R}^l$). The objective of LPP method is to minimize the distance of the nearest neighbour samples after projection. The objective function is as in (5)

$$\min \sum\nolimits_{i,j} \left\| z_i - z_j \right\|^2 W_{ij} \tag{5}$$

where W_{ij} is the weight of the edge connecting x_i and x_j. The weight W_{ij} is calculated by the heat kernel function as in (6):

$$W_{ij} = \begin{cases} \exp\left(-\dfrac{d_{ij}^2}{2\sigma^2}\right) \boldsymbol{x}_i \in \mathcal{N}_k\left(\boldsymbol{x}_j\right) \vee \boldsymbol{x}_j \in \mathcal{N}_k\left(\boldsymbol{x}_i\right) \\ 0 \qquad\qquad otherwise \end{cases} \tag{6}$$

where $\mathcal{N}_k\left(\boldsymbol{x}_j\right)$ refers to the k-nearest neighbour with the same label as $\boldsymbol{x}_j$. σ is a parameter to a heat kernel function. d_{ij} is the distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

It is worth noting that, if the range of the required transformation matrix A is not limited, then there is no solution to the minimization problem. Therefore, constraint $A^T X D X^T A = I$ is added. The final optimization problem of LPP method is as in (7):

$$\begin{aligned} &\arg\min_A \ tr\left(A^T X L X^T A\right) \\ &s.t.\ A^T X D X^T A = I \end{aligned} \tag{7}$$

When extended to the nonlinear case, the map $\phi(\boldsymbol{x}) : \mathbb{R}^d \to \mathcal{H}$ extends the data in Euclidean space $\mathbb{R}^d$ to Hilbert space $\mathcal{H}$. Then, the optimization problem becomes the form of (8), and v is the required transformation matrix:

$$\begin{aligned} &\arg\min_v tr\left(v^T \phi(X) L \phi(X)^T v\right) \\ &s.t.\ v^T \phi(X) D \phi(X)^T v = I \end{aligned} \tag{8}$$

where, $\phi(X) = [\phi(\boldsymbol{x}_1), \ldots, \phi(\boldsymbol{x}_{n+m})]$ represents the data in Hilbert space, and the Lagrange multiplier method is used to solve the above optimization problem, which can be transformed into solving the generalized eigenvalues, as shown in (9):

$$\left[\phi(X) L \phi(X)^T\right] v = \lambda \left[\phi(X) D \phi(X)^T\right] v \tag{9}$$

λ is Lagrange multiplier. According to [17], the eigenvector v is a linear combination of data $\phi(\boldsymbol{x}_1), \ldots, \phi(\boldsymbol{x}_{n+m})$. Then, there is coefficient matrix $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_{n+m}]^T \in \mathbb{R}^{(n+m)\times l}$, which make $v^T = [\phi(\boldsymbol{x}_1), \ldots, \phi(\boldsymbol{x}_{n+m})]\alpha$. Equation (8) can be rewritten to (10):

$$\begin{aligned} &\arg\min_\alpha tr\left(\alpha^T K L K \alpha\right) \\ &s.t.\ \alpha^T K D K \alpha = I \end{aligned} \tag{10}$$

where $K \in \mathbb{R}^{(n+m)\times(n+m)}$ is kernel matrix $K_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left\langle \phi(\boldsymbol{x}_i), \phi\left(\boldsymbol{x}_j\right)\right\rangle = \phi(\boldsymbol{x}_i)^T \phi\left(\boldsymbol{x}_j\right)$.

Since the maximum mean difference (MMD) technique used in the model measures the distance in Hilbert space, $\mathcal{L}_{LPP}$ is constructed based on the kernel LPP technique instead of the base LPP. $\mathcal{L}_{LPP}$ can be expressed directly as in (10). It is worth mentioning that the weight index is cosine distance rather than Euclidean distance due to that cosine distance can better describe the adjacent relation of vectors.

The definition of cosine similarity is expressed in (11).

$$d_{ij} = 1 - \frac{\phi\left(X_i\right), \phi\left(X_j\right)}{\left\|X_i\right\|_2 \left\|X_j\right\|_2} \tag{11}$$

Finally, $\mathcal{L}_{LPP}$ minimizes the angle of samples with the same label in the training data to achieve clustering.

(2) Description of $\mathcal{L}_{MMD}$ term

In order to increase the cluster ability, $\mathcal{L}_{MMD}$ is utilized to reduce the distance between data of source domain and target domain in training set. MMD estimates the difference between two distributions in a regenerated kernel Hilbert space (RKHS) [18]. $X_s = \{\boldsymbol{x}_{si}\}_{i=1}^n$ and $X_t = \{\boldsymbol{x}_{ti}\}_{i=1}^n$ are data sets of source domain and target domain respectively, and the distance between them is expressed as in (12):

$$MMD^2\left(X_s, X_t\right) = \left\| \frac{1}{n}\sum\nolimits_{i=1}^{n} \phi\left(\boldsymbol{x}_{si}\right) - \frac{1}{m}\sum\nolimits_{i=1}^{m} \phi\left(\boldsymbol{x}_{ti}\right) \right\|_{\boldsymbol{\mathcal{H}}}^2 \tag{12}$$

Let $\left\{C_1, C_2, \cdots, C_g\right\} \in \{1, \cdots, \mathrm{C}\}$ be the same fault label in both the source and target domains. $\mathcal{L}_{MMD}$ is constructed as in (13) and in (14).

$$\mathcal{L}_{MMD} = \sum\nolimits_{c=C_1}^{C_g} \mathcal{L}_{MMD}^c\left(\mathcal{Q}_s, \mathcal{Q}_t\right) \tag{13}$$

$$\mathcal{L}_{MMD}^c\left(\mathcal{Q}_s, \mathcal{Q}_t\right) = \left\| \frac{1}{n^c}\sum\nolimits_{\boldsymbol{x}_i \in \mathcal{D}_s^c} z\left(\boldsymbol{x}_i\right) - \frac{1}{m^c}\sum\nolimits_{\boldsymbol{x}_j \in \mathcal{D}_t^c} z\left(\boldsymbol{x}_j\right) \right\|_{\mathcal{H}}^2 \tag{14}$$

where $\mathcal{D}_s^c$ is the data of source domain that belongs to the label c. $\mathcal{D}_t^c$ is the data of target domain that belongs to label c.

As same as constructing $\mathcal{L}_{LPP}$, then substitute $z = v^T\phi\left(\boldsymbol{x}\right), v^T = [\phi\left(\boldsymbol{x}_1\right), \ldots, \phi\left(\boldsymbol{x}_{n+m}\right)]$. $\mathcal{L}_{MMD}$ can be rewritten as follows:

$$\mathcal{L}_{MMD} = \sum\nolimits_{c=C_1}^{C_g} tr\left(\alpha^T K M_c K \alpha\right) = tr\left(\alpha^T K M K \alpha\right) \tag{15}$$

where $M = \sum_{c-C_1}^{C_g} M_c$, and M_c can be calculated as in (16):

$$\left(M_c\right)_{ij} = \begin{cases} \dfrac{1}{n^c n^c} & \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{D}_s^c \\ \dfrac{1}{m^c m^c} & \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{D}_t^c \\ \dfrac{-1}{n^c m^c} & \boldsymbol{x}_i \in \mathcal{D}_t^c, \boldsymbol{x}_j \in \mathcal{D}_s^c \ or \ \boldsymbol{x}_i \in \mathcal{D}_s^c, \boldsymbol{x}_j \in \mathcal{D}_t^c \end{cases} \tag{16}$$

(3) Description of $\mathcal{L}_R$ term

The regularization term $\mathcal{L}_R$ can avoids the overfitting of this model. $\mathcal{L}_R$ can be described as in (17).

$$\mathcal{L}_R = tr\left(\alpha^T \alpha\right) \tag{17}$$

Substitute (10), (15), (17) into (4). The optimization problem of the final model can be written as (18):

$$\begin{aligned} &arg\min\ tr\left(\alpha^T K L K \alpha\right) + \lambda tr\left(\alpha^T K M K \alpha\right) + \mu tr\left(\alpha^T \alpha\right) \\ &s.t.\ \alpha^T K D K \alpha = I \end{aligned} \tag{18}$$

Solving (18) by Lagrange multiplier method, then this problem can be transformed into solving (19).

$$\max_{\alpha} tr\left[\left(\alpha^T\left(KLK + \lambda KMK + \mu I\right)\alpha\right)^{-1} \alpha^T K D K \alpha\right] \tag{19}$$

The solution of the above problem is similar to the TCA method [19] and Fisher discriminant analysis [20]. The final solution α_* is a matrix of the eigenvectors corresponding to the largest l eigenvalues of $(KLK + \lambda KMK + \mu I)^{-1} KDK$.

Project the MFE features of the source and target domains into the l-dimensional subspace as in (20) and in (21), then the MTFE feature can be obtained:

$$z\left(X_t\right) = v^T\phi\left(X_t\right) = \alpha_*^T K\left(X_s, X_t\right) \tag{20}$$

$$z\left(X_s\right) = v^T\phi\left(X_s\right) = \alpha_*^T K\left(X_s, X_s\right) \tag{21}$$

Six parameters appeared the transfer learning model.

(1) σ: Scale parameter of heat kernel $\sigma \in \mathbb{R}$ influences the value of the similarity matrix W.

(2) k: The number of nearest neighbors' nodes $k \in \mathbb{N}$ controls the construction of the adjacency graph.

(3) λ: MMD trade-off parameter $\lambda > 0$ affects the degree of penalty in reducing the distribution discrepancy between normal samples of different domains.

(4) μ: Regularization parameter $\mu > 0$ controls model complexity.

(5) l: Dimensionality of subspace $l > 0$ affect the dimension of MTFE features.

(6) τ: The scale factor τ of the original feature determines the dimension of the input feature.

As for how to select these parameters for our model, there are two parts need to be considered. First, the parameters are adjusted iteratively through the data of the training set. In transfer learning scenario, the data distribution of the source domain and the target domain is different, so the cross-validation strategy of traditional classification algorithms cannot be used. When using the MTFE algorithm, the model is trained using source domain data, but the accuracy of diagnosis under different parameter combinations is tested using target domain data instead of source domain data. Then the parameter combination is constantly adjusted until the highest accuracy is reached. It can reduce the risk of overfitting the model to the source domain data and improves the generalization of this model.

After that, a more efficient search algorithm is needed to choose different combinations of parameters. In the MTSDE algorithm, bayesian optimization algorithm is used instead of the traditional grid search or random search algorithm. There are three main reasons about why bayesian optimization is chosen instead of traditional grid and random search methods:

(1) Bayesian optimization algorithm utilizes the gauss process to constantly update the a priori parameter information, while other search methods cannot take into account the previous parameter information.

(2) Bayesian optimization requires fewer iterations when searching for the optimal parameter combinations so that it is

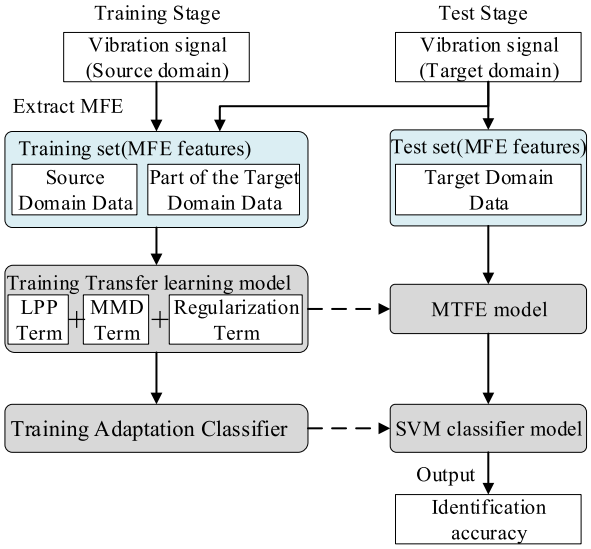

FIGURE 3. Flowchart of MTFE-SVM cross-domain intelligent fault diagnosis method.

in high calculation efficiency. By contrast, the grid search is not effective in optimizing many parameters.

(3) Bayesian optimization performs better than the grid search method for the non-convex problems because the grid search method is easy to get local optimal value.

C. STEPS OF PROPOSED MTFE METHOD

The last part of this chapter summarizes the steps of MTFE method.

Step 1: Extract the MFE features of vibration signals in both the target domain and source domain, expressed as the matrix $\left[\boldsymbol{x}_{s1}^{d}, \boldsymbol{x}_{s2}^{d}, \ldots, \boldsymbol{x}_{sn}^{d}\right]$ and $\left[\boldsymbol{x}_{t1}^{d}, \boldsymbol{x}_{t2}^{d}, \ldots, \boldsymbol{x}_{tm}^{d}\right]$.

Step 2: Input the features of training set (In this paper, it consists of the source domain and a part of the normal sample of the target domain) into the transfer model to obtain the projection matrix $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_{n+m}]^T \in \mathbb{R}^{(n+m)\times l}$, where l is the dimension of MTFE features.

Step 3: Bayesian optimizer is used to further optimize the model hyper parameters in order to obtain the optimal MTFE features, including the number of nearest neighbour node k, MMD trade off parameter λ, regularization parameter μ, the dimensions of the subspace l and scale factor τ. In addition, penalty coefficient c and kernel radius g of SVM are optimized at the same time. Output projection matrix α.

Step 4: Project the MFE features of training set and test set into the l-dimensional subspace using α. The low dimensional data after projection is the MTFE feature.

Step5: Input the MTFE features into SVM classifier for training and testing.

The flowchart of MTFE-SVM cross-domain intelligent fault diagnosis method is given in Figure. 3.

IV. EXPERIMENT

In this section, fault diagnosis experiments on rolling bearings and gearboxes are designed to verify the validity of the MTFE algorithm.

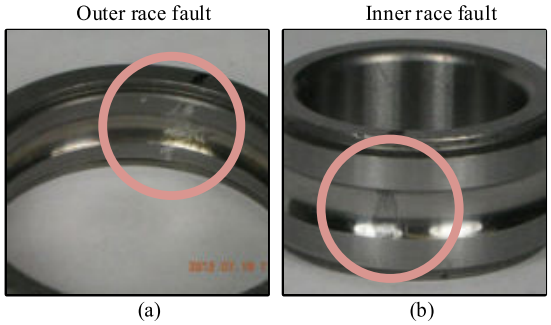

FIGURE 4. (a) Outer race fault of bearing, (b) Inner race fault of bearing.

A. DESCRIPTION OF DATASETS

1) CWRU (BEARING)

This dataset is collected from bearing data centre of Case Western Reserve University (CWRU) [27]. The platform consists of four parts, a 1.5kW (2hp) electric motor, a torque sensor/decoder, a power meter and electronic controller. The dataset contains acceleration signals collected from two bearings (drive end SKF 6205rs with sampling frequency 12kHz, fan end SKF 6203rs with sampling frequency 12kHz and 48kHz) with four health states (normal, outer race fault, inner race fault and ball fault, each fault type contains three types of fault severity) under four motor loads (0,1,2 and 3hp).

2) MFPT (BEARING)

This dataset is provided by Society for Machinery Failure Prevention Technology (MFPT) [28]. The test rig was equipped with a rolling bearing manufactured by NICE (roller diameter: 0.235, pitch diameter: 1.245, number of elements: 8, contact angle 0°). This data set consists of three normal conditions (collected under 270lbs of load, input shaft rate of 1500rpm and sampling rate of 97,656sps), three outer ring fault data collected under 270lbs of load, input shaft rate of 1500rpm and sampling rate of 97,656sps), another seven outer ring fault data (collected under 25, 50, 100, 150, 200, 250, 300lbs of load, input shaft rate of 1500rpm and sampling rate of 48828sps) and seven inner ring fault data (collected under 0, 50, 100, 150, 200, 250, 300lbs of load, input shaft rate of 1500rpm and sampling rate of 48828sps). The bearings of the two fault types are shown in Figure. 4. The test rig was equipped with a NICE bearing with the following parameters: roller diameter: rd = 0.235, pitch diameter: pd = 1.245, number of elements: ne = 8 and contact angle: ca = 0°. For details on experiment design and the experiments conducted, readers are referred to the Ref. [28].

3) PHM09 (GEARBOX)

This dataset is provided from the Prediction and Health Management Society for the 2009 data challenge (PHM 2009 Data Challenge) [29]. Schematic of the gearbox used for vibration data acquisition is shown in Figure.4. The gearbox contains three shafts, four gears (input pinion: 16T or 32T; idler gear I: 48T or 96T; idler gear II: 24T or 48T; output gear: 40T or 80T)

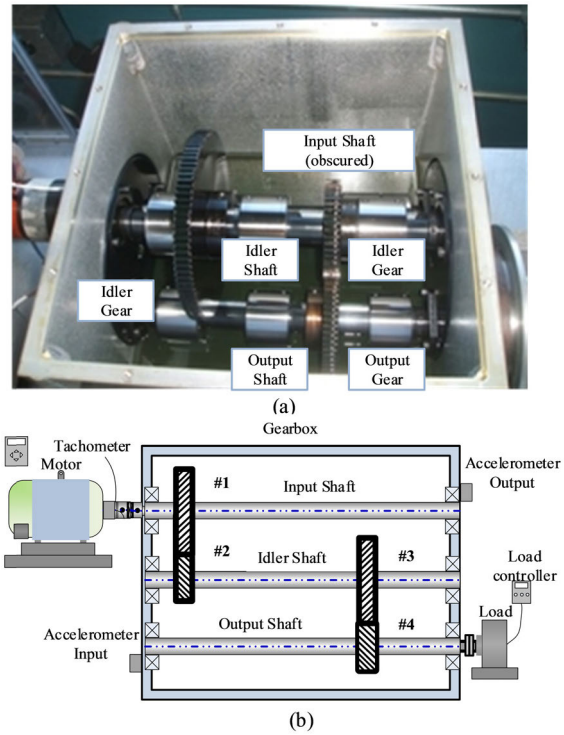

FIGURE 5. (a) Inside view of gearbox used for data collection and testing (PHM 2009 Data Challenge), (b) Schematic of gearbox test rig used in PHM 2009 Challenge Data.

and six bearings (roller diameter: rd = 0.3125mm 4; pitch diameter: pd = 1.319; number of elements: ne = 8; contact angle: ca = 0). The vibration data collected from three sensors, two accelerometers and one tachometer. The complete dataset consisted of 560 samples, in which gearboxes were tested under five different speeds (30, 35, 40, 45, 50Hz), two different loads (heavy and light) and two different gear types (spur and helical), including eight health condition of spur gears and six of helical gears. The sampling frequency is 66.67kHz. For details on experiment design and the experiments conducted, readers are referred to the Ref. [29].

B. EXPERIMENTAL SETUP

Experiment I : A cross-domain rolling bearing diagnosis experiment is designed. A/B/C is used to represent three domains respectively. The composition of each domain is shown in the Table 1. In order to verify the generalization ability of the proposed method, six diagnosis tasks A-B, A-C, B-A, B-C, C-A, C-B were designed, where Diagnosis tasks A-B means that domain A is used as the source domain to identify faults types in target domain B. Half of the normal target domain samples and the source domain constitute the training set, and the target domain data without labels constitute the test set. Note that the ball fault samples were removed from the A-C, B-C, C-A, C-B experiments, because domain C does not have fault samples of this type.

Experiment II: In order to verify the performance of MTFE method in rotating machinery fault diagnosis, a diagnosis experiment of gearbox under different working conditions

TABLE 1. The composition of the domain in Experiment I.

Domain	Dataset	Condition	Label	Fault Category	Number
A	CWRU	29.95 Hz (1797 rpm) 0hp 3 fault levels	1 2 3 4	N OR IR B	90 3×30=90 3×30=90 3×30=90
B	CWRU	28.83 Hz (1730 rpm) 3hp 3 fault levels	1 2 3 4	N OR IR B	90 3×30=90 3×30=90 3×30=90
C	MFPT	25Hz-270lbs 25Hz-270lbs 25Hz-250/300lbs	1 2 3	N OR IR	100 100 100

TABLE 2. The composition of the domain in Experiment II.

Domain	Dataset	Condition	Label	Fault Category	Number
Gear-30L [35L/40L/ 45L/50L]	PHM09	30 Hz [35 Hz/40 Hz/45 Hz/50 Hz] Low load	1 2 3	N CT MT	100+100 100 100
Gear-30H [35H/40H/ 45H/50H]	PHM09	30 Hz [35 Hz/40 Hz/45 Hz/50 Hz] High load	1 2 3	N CT MT	100+100 100 100

is designed. The experimental set consists of vibration data of helical gear in three health states (normal, chipped tooth, and missing tooth) in PHM09 data set and composition of each domain is shown in the Table. Based on these domains, six gearbox fault diagnosis tasks under different working conditions are organized. Two tasks were performed under different shaft speeds and loads (30L-40H, 45H-50L), two tasks were performed under different shaft speeds (30H-50H, 40L-45L), and three tasks were performed under different loading conditions (40L-40H, 45H-45L). 30H in the diagnosis task represents the data collected under heavy load with the shaft speed of 30Hz [25]. The composition of the training set is the same as that of Experiment I, and the remaining target domain data without labels constitute the test set.

C. EXPERIMENTAL PROCESS

In order to show the superiority of MTFE-SVM intelligent fault diagnosis method proposed in this paper, three traditional classifier models and three transfer learning methods are set as control group.

When implementing the following five methods, the MFE features of the training set and the test set are extracted first. Embedding dimension $\tilde{m}$, similar tolerance r, scale factor τ is set to 2, 0.15 and 25, respectively.

1) MFE-SVM

Support vector machine (SVM) classifier is trained using the MFE of training set directly, namely MFE-SVM. Then, the identification accuracy of the test set is output. The LIBSVM toolbox is used to implement SVM. The RBF kernel function is applied and trade off parameter c is set to 1.

TABLE 3. Identification accuracy of Experiment I.

Experiment I	MFE-SVM	MFE-KNN	MFE-LR	MFE-GFK-SVM	MFE-DAFD-SVM	MFE-SSTCA-SVM	MTFE-SVM
A-B	89.4%	**93.6%**	79.2%	78.1%	38.8%	80.4%	90.8%
B-A	90.3%	**93.6%**	76.1%	83.9%	35.8%	63.9%	91.4%
A-C	85.7%	77.0%	60.7%	53.0%	35.3%	77.0%	**89.0%**
C-A	50.4%	71.5%	54.8%	33.3%	36.1%	72.6%	**75.6%**
B-C	89.0%	**89.0%**	63.0%	67.0%	35.6%	75.3%	**89.0%**
C-B	55.2%	66.7%	53.7%	31.5%	35.1%	**70.7%**	64.8%
Average	76.7%	81.9%	64.6%	57.8%	36.1%	73.3%	**83.4%**

2) MFE-KNN

The experimental procedure is the same as MTFE-SVM, except that k-nearest neighbour (KNN) classifier is used instead of SVM. The optimal number of nearest neighbour nodes is searched in {1, 5, 9, 13, 17, 21, 25, 29, 33, 63}, then the result corresponding to the optimal parameter is output.

3) MFE-LR

Like (1) and (2), logistic regression (LR) classification method is applied and the best trade off parameter is searched in {0.001, 0.01, 0.1, 1, 10}. Then the best accuracy is output.

4) MFE-GFK-SVM

For MFE-GFK, first, the MFE feature is extracted, then the feature is mapped by geodesic flow kernel (GFK) method [30], and finally the SVM classifier is used to classify it. The subspace embedding approach is the principal component analysis.

5) MFE-SSTCA-SVM

Input the MFE features into the semi-supervised transfer component analysis (SSTCA) [21] model. The optimal hyper-parameters are searched by Bayesian optimization method. Search MFE features scale factor τ, regularization trade-off parameter μ_1, subspace dimension l_1, supervise term trade-off parameters γ_1 and geometry term trade-off parameters λ_1 in the range [1, 25] (in Experiment II, [1, 50]), [10^{-3},10^3], [1, 10], [10^{-3}, 1] and [10^{-3},10^3] respectively. Finally input the low-dimensional features into SVM classifier and output the diagnosis accuracy.

6) MFE-DAFD-SVM

For this approach, take Deep neural network for domain in fault diagnosis (DAFD) as the transfer strategy [31]. The MFE feature is used as input to a neural network with three layers, and then the output feature is classified using the SVM classifier.

7) MTFE-SVM

MTFE-SVM intelligent fault diagnosis model is applied according to steps from Chapter III.C. For this method, the heat kernel parameter σ is set to constant as 1. Then Bayesian optimization algorithm is applied to search for the optimal parameters of MTFE-SVM method in the hyper-parameter space. Search optimal parameters τ, k, l, λ, μ in [1,25] (in Experiment II, [1,50]), [1,11], [1,10], [10^{-3},10^3], and [10^{-3},10^3]. Moreover, the two parameters c and g appeared in SVM are also optimized using Bayesian method. Search them in (0, 8] and finally get the accuracy.

D. EXPERIMENT RESULT ANALYSIS

1) IDENTIFICATION ACCURACY

The fault diagnosis accuracy of Experiment I is shown in the TABLE 3. It can be seen that in the three traditional machine learning methods, the accuracy of MFE-KNN reaches 81.9%, which is much higher than the others. Especially, the accuracy rate is the highest in task A-B and B-A. This is because domain A and domain B are so similar that diagnosis can be achieved using traditional classification algorithms, while transfer learning algorithms may be negative. All three transfer learning methods perform poorly. The algorithms MFE-GFK-SVM and MFE-DAFD-SVM are difficult to identify the category of fault. The transfer learning method MFE-SSTCA-SVM performs better, but it is still worse than traditional MFE-KNN and MFE-SVM algorithms. The possible reason is that the main strategy of SSTCA is to reduce the maximum mean difference between target domain and source domain data, while the training set lacks fault samples of target domain, leading to its performance degradation.

Obviously, the MTFE-SVM method proposed in this paper achieved the highest mean accuracy of 83.4%, which proved that the MTFE features indeed preserved the similar spatial structure of the two domains and reduced the distribution distance.

As shown in TABLE 4, the above results are clearer in the fault diagnosis experiment of more complicated rotating mechanic, gearbox. In the fault diagnosis experiments on gearboxes, the traditional classification algorithms MFE-SVM, MFE-KNN, and MFE-LR all perform worse because gearbox vibrations are more complex, resulting in greater distribution differences between features. Transfer learning algorithms show advantages in this experiment. Especially, the results of MFE-SSTCA-SVM method considering data transfer are significantly better than the three traditional machine learning methods without transfer, while the other two transfer learning algorithms are also lower than it.

However, the accuracy of MTFE-SVM is up to 90.2%, 17.2% higher than MFE-SSTCA-SVM. This is because the MTFE-SVM algorithm is more comprehensive in that it not

TABLE 4. Identification accuracy of Experiment II.

Experiment II	MFE-SVM	MFE-KNN	MFE-LR	MFE-GFK-SVM	MFE-DAFD-SVM	MFE-SSTCA-SVM	MTFE-SVM
30L-40H	40.0%	38.3%	71.0%	50.0%	33.3%	60.3%	**85.0%**
45H-50L	33.3%	34.0%	46.0%	36.3%	33.3%	86.7%	**93.3%**
45H-45L	37.0%	41.7%	44.3%	45.3%	33.3%	73.3%	**99.0%**
35L-35H	38.3%	37.7%	64.0%	41.0%	33.3%	60.3%	**82.0%**
40L-45L	57.7%	64.7%	75.0%	65.7%	34.9%	87.0%	**99.3%**
30H-50H	55.7%	66.3%	73.7%	66.0%	33.3%	70.3%	**82.7%**
Average	43.7%	47.1%	62.3%	50.7%	33.6%	73.0%	**90.2%**

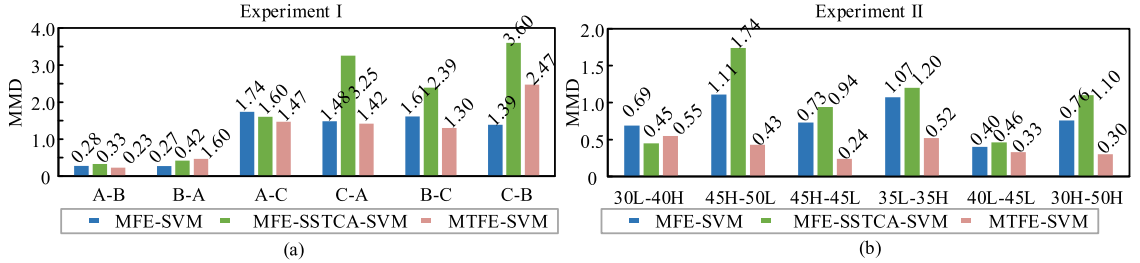

FIGURE 6. Histograms of MMD (a) Experiment I, (b) Experiment II.

only reduces the maximum mean difference between the source and target domains, but also considers the spatial distribution structure of the data, which helps to achieve classification.

2) MAXIMUM MEAN DIFFERENCE BETWEEN TARGET DOMAIN AND SOURCE DOMAIN

According to [23], maximum mean difference (MMD) can be used as an index to measures the distribution difference between source domain and target domain. Therefore, the diagnostic accuracy is negatively correlated with MMD. In order to illustrate the effectiveness of the method proposed in this paper, the MMD values of the source and target domains in each diagnostic task in the original eigenspace, the SSTCA subspace, and the MTFE subspace are shown in Figure.6.

In the experiment, parameters of SSTCA and MTFE model are set as the optimal parameters obtained by Bayesian optimization. Note that the features dimensions in subspace are set to 2, which is no need for optimization. In addition, since the MMD value is related to the feature dimension, the scale parameter of the RBF kernel in the original MFE space is set as 3.53 (5 in Experiment II) to eliminate this effect.

It can be observed from the Figure.6 that in Experiment I, the MMD value of the MTFE-SVM method is minimized in the four diagnostic tasks (A-B, A-C, C-A and B-C). In Experiment II, the MMD values of six diagnostic tasks are all smaller than the other two methods. This result is basically consistent with the diagnosis accuracy, and the task accuracy with low MMD value is correspondingly higher. It is proved that the MTFE-SVM algorithm proposed in this paper can effectively reduce the difference of feature distribution in different domains.

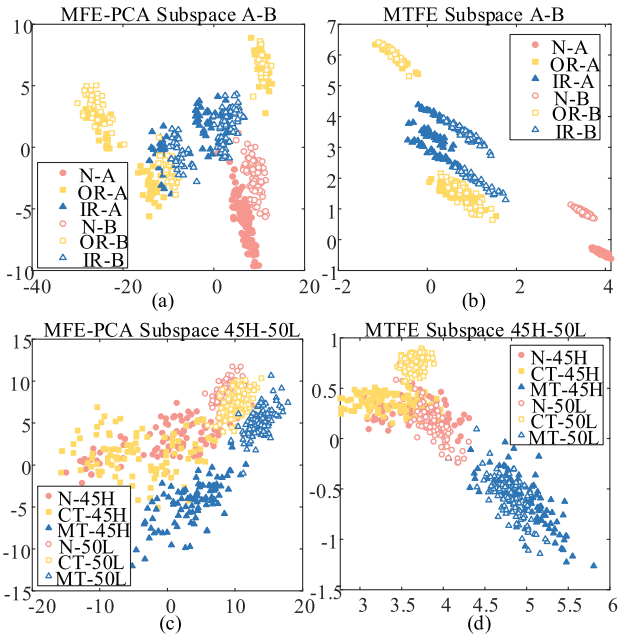

FIGURE 7. Feature visualization (a) Task A-B in PCA subspace, (b) Task A-B in MTFE subspace, (c) Task 45H-50L in PCA subspace, (d) Task 45H-50L in MTFE subspace.

3) FEATURE VISUALIZATION

In order to demonstrate the superiority of MTFE features intuitively, the original MFE features and MTFE features are projected into a two-dimensional plane respectively, as shown in Figure.7. The Figureure shows one diagnostic task in each of the two experiments (A-B and 45H-50L). The principal component analysis (PCA) method is used to reduce dimension of original MFE features to 2. For the MTFE method, parameter l is set as constant 2 while other parameters need to be optimized.

TABLE 5. Identification accuracy of different entropy.

	A-B	B-A	A-C	C-A	B-C	C-B	Average
MTSE-SVM	86.1%	90.6%	87.0%	68.9%	66.7%	**78.1%**	79.6%
MTPE-SVM	79.4%	73.6%	62.0%	65.6%	66.7%	68.5%	69.0%
MTFE-SVM	**90.8%**	**91.4%**	**89.0%**	**75.6%**	**89.0%**	64.8%	**83.4%**

In the rolling bearing diagnosis task (shown in Figure.7(a) and (b)), it is difficult to distinguish the inner ring fault class from the outer ring fault class in PCA subspace of MFE due to the overlap. However, there is almost no overlap between the three classes in MTFE subspace. Similarly, normal samples and chipped tooth samples are indistinguishable in the subspaces of MFE in gearbox diagnostic experiments, but the two categories are more separable in MTFE subspace. The phenomenon shown in Figure.8 effectively proves that the MTFE feature is helpful to improve the classification effect and can identify the fault type more accurately.

4) COMPARISON WITH OTHER ENTROPY-BASED METHOD

Taking Experiment I as an example, we added comparisons with other entropy algorithms, including multi-scale sample entropy and multi-scale permutation entropy. We supplemented the same transfer strategy, only the input features are different, so called them MTSE-SVM and MTPE-SVM.

The results show that the MTFE-SVM algorithm obtains the highest accuracy in all tasks except diagnostic task A-B. Because our transfer strategy is shallow, the accuracy of feature extraction greatly affects the final diagnosis. The MSE feature uses only the Heaviside function to determine the similarity of the two signals, which is less effective than the fuzzy function used by MFE. MPE reflects system changes by symbolizing time series, and although it achieves high computational efficiency, it is less accurate than MFE. Therefore, we believe that MTFE features that extend based on MFE are more effective.

5) COMPARISON WITH DEEP DOMAIN ADAPTION METHODS

Transfer learning algorithms based on artificial intelligence have been widely used, therefore, the performance of several deep domain adaptation methods on bearing diagnosis tasks are shown here. Three deep domain adaptation methods are considered in this paper. They are the remarkable representations of deep domain adaptation methods.

(1) DANN: domain adversarial neural network [32]

(2) CDAN: conditional domain adversarial network [33]

(3) MK-MMD: multi kernel maximum mean discrepancy [34]

Please note that the detailed parameters setting of above three methods are following [35]. For these methods, the original vibration signals are fed into the networks directly.

The performance of these three deep domain adaptation methods is considered, their diagnosis results are listed in Table 6. As can be seen, the obtained results of the three deep learning methods are not satisfactory. This can be explained in the following way. The experimental conditions used for validation is exactly harshly. The source and target domains are collected from completely different data sets, operating conditions, bearing models, signal sample rates, and data length, etc., which will cause a serious impact on the classification performance of deep learning domain adaption methods. Unlike these deep learning methods, MTFE algorithm is firstly utilized to extract the fault features from the vibration signal, thereby, the mentioned difference between the source and target domains can be significantly reduced. Above all, the generality of MTFE method is better than that of deep domain adaption methods.

TABLE 6. Identification accuracy of deep domain adaptions.

	CNN-DANN	CNN-CDAN	CNN-MK-MMD	MTFE-SVM
A→B	91.40%	86.10%	**94.40%**	90.80%
B→A	**98.60%**	94.40%	91.70%	91.40%
A→C	25.00%	25.00%	25.00%	**89.00%**
C→A	25.00%	25.00%	25.00%	**75.60%**
B→C	25.00%	25.00%	25.00%	**89.00%**
C→B	25.00%	25.00%	25.00%	**64.80%**
Average	48.30%	46.80%	47.70%	**83.40%**

V. CONCLUSION

In this paper, the MTFE-SVM intelligent fault diagnosis method is proposed for rotating machinery fault diagnosis. Firstly, MFE features are extracted from training samples. Secondly, the MTFE features are obtained by reducing dimensions of MFE features, in which the MMD between training data and test data is reduced. Finally, the MTFE features are as input and the SVM classifier is used to identifiers the fault type. The main contributions of this paper include:

(1) The MTFE is an effectively features extraction method used for fault diagnosis. Compared with the MFE, MTFE can not only effectively extract features that is easily to identify faults, but also can significantly minimized the distribution discrepancy between source domain and target domain data. In the fault diagnosis experiments of bearings and gearboxes, the accuracy of MTFE-SVM algorithm improved by 6.7% and 46.5%, respectively, compared to traditional MFE-SVM. It is suitable for fault diagnosis task under cross-domain.

(2) A novel cross-domain fault diagnosis method called MTFE-SVM is proposed and the advantages of this method are verified using two experiments. As shown in result, MTFE-SVM has an accuracy rate of 83.4% and 90.2%, higher than other classification algorithms. Compared with other methods, this method has a good generalization ability and higher accuracy under different domain data collected both in same machine and in different machine.

The proposed method provides a new idea for RMFD in the absence of target machine data. In fact, in industrial

production, it is difficult to collect abundant amounts of data on equipment that needs to be diagnosed. This is a test for intelligent fault diagnosis methods. The MTFE-SVM method proposed in this paper can realize the identification of the fault location of the equipment by the model trained by the data collected from other similar fault equipment or simulation equipment in the laboratory.

In future work, we will test the effectiveness of MTFE in feature extraction of other types of signals, such as acoustic signal, electrocardiograph (ECG), and digital images. In addition, the effectiveness of multiple-fault diagnosis is unknown. Further test using multiple fault of rotating machinery will be considered in our future work.

REFERENCES

[1] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587.

[2] S. Sahin, M. R. Tolun, and R. Hassanpour, "Hybrid expert systems: A survey of current approaches and applications," *Expert Syst. Appl.*, vol. 39, no. 4, pp. 4609–4617, Mar. 2012.

[3] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, Aug. 2007.

[4] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mech. Syst. Signal Process.*, vol. 107, pp. 241–265, Jul. 2018.

[5] D. Goyal, S. S. Dhami, and B. S. Pabla, "Non-contact fault diagnosis of bearings in machine learning environment," *IEEE Sensors J.*, vol. 20, no. 9, pp. 4816–4823, May 2020.

[6] Y. Huang, D. Yang, K. Wang, L. Wang, and Q. Zhou, "Stability analysis of GMAW based on multi-scale entropy and genetic optimized support vector machine," *Measurement*, vol. 151, Feb. 2020, Art. no. 107282.

[7] H. Liu, C. Liu, and Y. Huang, "Adaptive feature extraction using sparse coding for machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 25, no. 2, pp. 558–574, Feb. 2011.

[8] Y. Li, X. Wang, Z. Liu, X. Liang, and S. Si, "The entropy algorithm and its variants in the fault diagnosis of rotating machinery: A review," *IEEE Access*, vol. 6, pp. 66723–66741, 2018.

[9] Y. Li, F. Liu, S. Wang, and J. Yin, "Multiscale symbolic Lempel–ziv: An effective feature extraction approach for fault diagnosis of railway vehicle systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 199–208, Jan. 2021.

[10] A.-S. Qin, H.-L. Mao, and Q. Hu, "Cross-domain fault diagnosis of rolling bearing using similar features-based transfer approach," *Measurement*, vol. 172, Feb. 2021, Art. no. 108900.

[11] C. Yang and M. Jia, "Hierarchical multiscale permutation entropy-based feature extraction and fuzzy support tensor machine with pinball loss for bearing fault identification," *Mech. Syst. Signal Process.*, vol. 149, Feb. 2021, Art. no. 107182.

[12] Wei, Li, Xu, and Huang, "A review of early fault diagnosis approaches and their applications in rotating machinery," *Entropy*, vol. 21, no. 4, p. 409, Apr. 2019.

[13] A. N. Kolmogorov, "A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces," *Doklady Akademii Nauk SSSR*, vol. 119, pp. 861–864, Jan. 1958.

[14] S. Pincus, "Approximate entropy (ApEn) as a complexity measure," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 5, no. 1, pp. 110–117, Mar. 1995.

[15] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol.-Heart Circulat. Physiol.*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000.

[16] W. Chen, J. Zhuang, W. Yu, and Z. Wang, "Measuring complexity using FuzzyEn, ApEn, and SampEn," *Med. Eng. Phys.*, vol. 31, no. 1, pp. 61–68, Jan. 2009.

[17] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, Apr. 2002, Art. no. 174102.

[18] Y. Li, Y. Yang, G. Li, M. Xu, and W. Huang, "A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection," *Mech. Syst. Signal Process.*, vol. 91, pp. 295–312, Jul. 2017.

[19] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, Jul. 2002, Art. no. 068102.

[20] J. Zheng, J. Cheng, Y. Yang, and S. Luo, "A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination," *Mechanism Mach. Theory*, vol. 78, pp. 187–200, Aug. 2014.

[21] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," in *Proc. 23rd Nat. Conf. Artif. Intell.*, vol. 2, Chicago, IL, USA, 2008, pp. 677–682.

[22] Y. Wei, Y. Yang, M. Xu, and W. Huang, "Intelligent fault diagnosis of planetary gearbox based on refined composite hierarchical fuzzy entropy and random forest," *ISA Trans.*, vol. 109, pp. 340–351, Mar. 2021.

[23] B. Quanz and J. Huan, "Large margin transductive transfer learning," in *Proc. 18th ACM Conf. Inf. Knowl. Manage.*, Hong Kong, 2009, pp. 1327–1336.

[24] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.

[25] H. Zheng, R. Wang, J. Yin, Y. Li, H. Lu, and M. Xu, "A new intelligent fault identification method based on transfer locality preserving projection for actual diagnosis scenario of rotating machinery," *Mech. Syst. Signal Process.*, vol. 135, Jan. 2020, Art. no. 106344.

[26] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[27] *Case Western Reserve University rolling bearing dataset*. [Online]. Available: https://csegroups.case.edu/bearingdatacenter/pages/12k-drive-end-bearing-fault-data

[28] *Eric Bechhoefer, MFPT Bearing Fault Data Sets*. [Online]. Available: http://mfpt.org/fault-data-sets/

[29] *PHM 09 Data Challenge Data*. [Online]. Available: https://www.phmsociety.org/competition/PHM/09/apparatus

[30] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2066–2073.

[31] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296–2305, Mar. 2017.

[32] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, 2016.

[33] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," 2017, *arXiv:1705.10667*. [Online]. Available: http://arxiv.org/abs/1705.10667

[34] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," 2015, *arXiv:1502.02791*. [Online]. Available: http://arxiv.org/abs/1502.02791

[35] Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, and X. Chen, "Unsupervised deep transfer learning for intelligent fault diagnosis: An open source and comparative study," 2019, *arXiv:1912.12528*. [Online]. Available: http://arxiv.org/abs/1912.12528

ZHENG DANGDANG was born in 1986. He is currently pursuing the degree with the Shaanxi Engineering Laboratory for Transmissions and Controls, Northwestern Polytechnical University (NWPU), China. He is also working with the First Aircraft Institute, AVIC. His research interests include model-based definition (MBD), model-based systems engineering (MBSE), and collaborative design and simulation.

BING HAN was born in 1981. He is currently an Engineer with the Shaanxi Engineering Laboratory for Transmissions and Controls, Northwestern Polytechnical University (NWPU), China. His research interests include collaborative design and simulation, integration technology of mechanical systems, and data and process management of design process

YONGBO LI (Member, IEEE) received the master's degree from Harbin Engineering University (HRBEU), Harbin, China, in 2012, and the Ph.D. degree in general mechanics from the Harbin Institute of Technology (HIT), Harbin, in 2017. He is currently an Associate Professor with the School of Aeronautics, Northwestern Polytechnical University, China. Prior to joining Northwestern Polytechnical University in 2017, he was a Visiting Student with the University of Alberta, Edmonton, AB, Canada. He also held five research grants from the National Natural Science Foundation of China, the Shaanxi Province Key Research Program, the China Postdoctoral Science Foundation, and so on. His research interests include signal processing, fault feature extraction, and fault pattern identification.

GENG LIU was born in 1961. He is currently a Professor, a Ph.D. Supervisor, and the Director of Shaanxi Engineering Laboratory for Transmissions and Controls, Northwestern Polytechnical University. His research interests include mechanical transmissions, dynamics of mechanical systems, virtual and physical prototyping simulation and design technology of mechanical systems, finite element methods, and contact mechanics.

HUANGCHAO YU received the Ph.D. degree in mechanical engineering from the University of Alberta, Canada, in 2017. He is currently working as an Associate Professor with the National University of Defense Technology. His research interests include solid mechanics, optimization, and intelligent design and control of unmanned systems.