RESEARCH ARTICLE

Research on a Multisource Domain Improved Fault Diagnosis Method of the Rotor System

YANG XIAO1,2, QINGFENG WANG1,2, SHUAI WANG1,2, AND WENWU CHEN3

1Beijing Key Laboratory of Health Monitoring and Self-Recovery of High-End Machinery Equipment, Beijing University of Chemical Technology, Beijing 100029, China
2School of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029, China
3State Key Laboratory of Safety and Control for Chemicals, China Petroleum & Chemical Corporation Research Institute of Safety Engineering, Qingdao 266071, China

Corresponding author: Qingfeng Wang (wangqf2422@buct.edu.cn)

This work was supported in part by the China Petroleum and Chemical Corporation Ministry of Science and Technology under Grant 321123-3, in part by the Open Fund of the State Key Laboratory of Compressor Technology under Grant SKL-YSJ202101, in part by the National Pipeline Network Group under Grant WZXGL202107, and in part by the Science and Technology Project of China National Petroleum Corporation under Grant 2021DJ6505.

ABSTRACT In this paper, rotor systems of the rotating machinery such as steam turbines, centrifugal compressors and flue gas turbines are selected as the research objects. At present, most of the rotor system fault diagnosis methods based on artificial intelligence algorithms are in the laboratory research stage, and there is still a gap from the actual industrial application. Therefore, the multi-source domain improved fault diagnosis (MSDIFD) method for satisfying engineering applications is proposed in this paper. Firstly, typical labeled data are selected to construct a multi-source domain training feature space. Then, commonality fault features are extracted and screened by the improved adaptive variational mode decomposition (IAVMD), and the feature reconstruction signal is automatically output. Next, the weighted semi-supervised transfer component analysis (WSSTCA) method based on enhanced kernel function is employed to narrow the disparity between feature vectors of cross domain data. Finally, typical failure case data and real-time monitoring data are used as the training data and test data of the model, respectively, and an ensemble fault recognition classifier is constructed to achieve failure mode identification of the rotor system. Using 40 groups of typical fault engineering cases under different equipment and different operating conditions, the proposed rotor fault identification method has been verified and compared with five published fault identification methods. The results indicate that the proposed method possesses more excellent fault diagnosis accuracy and domain generalization performance, and the MSDIFD method has good application and promotion value for solving cross equipment, cross working condition, and cross domain diagnostic tasks.

INDEX TERMS Multi-source domain improved fault diagnosis, domain generalization, rotor system, weighted semi-supervised transfer component analysis, improved adaptive variational mode decomposition.

I. INTRODUCTION

The rotor system is the core component of rotating machinery. Once it fails, it directly affects the working state of the entire rotating machinery. It might even cause shutdowns or equipment damage accidents [1]. Digital transformation of enterprises puts forward new requirements for predictive maintenance (PdM). Intelligent fault diagnosis of mechanical parts is the basis of predictive maintenance. Consequently, deep look into the fault diagnosis technique of rotor systems is all-important for achieving PdM, ensuring the safe operation of rotating machinery, and eliminating accidents.

There are many types of faults in the rotor system of rotating machinery. The most typical ones include: shaft misalignment, rotor unbalance, oil whirl and oil whip, rubbing, and surge [2]. Since many reasons may cause the same fault and many faults have some same symptoms, it is generally not suitable to disassemble rotating machinery and check
for faults [3]. The aforementioned characteristics increase the difficulty of state monitoring and failure analysis of the rotor system. The current research on failure recognition of rotor systems mainly focuses on the following aspects. Principles of rotor dynamics are used to research fault mechanisms [4], while time-frequency analysis, transient analysis, and other vibration signal analysis techniques are used for fault diagnosis [5]. Artificial intelligence expert systems and neural networks are used for fault diagnosis [6] as well as the development and research of fault diagnosis devices [7]. In recent years, pattern recognition theory has been applied in rotor system fault identification [8]. The rotor system fault recognition method is based on isometric feature mapping which evaluates shaft trajectory [9]. The rotor rubbing fault diagnosis method is a new frequency analysis method based on extracting the highly oscillatory frequency which calls matching demodulation transform [10]. On the other hand, rotor-shaft system fault identification is based on an improved stacked denoising autoencoder [11]. Another typical example is the method founded on direct connection-based convolutional neural network (DC-CNN) [12]. However, many rotor equipment condition monitoring systems and fault diagnosis methods often remain in the research stage. Their diagnostic accuracy and precision of practical engineering fault cases have not reached the ideal level. Consequently, these methods are ordinarily difficult to employ in the industrial environment.

Variational mode decomposition (VMD) is able to effectively decrease the noise component interference of rotor vibration signals in practical engineering [13]. Zhu and Xia [14] established a two-span rotor system to verify the effectiveness of the VMD method in identifying rubbing faults. However, this method without screening the sensitive intrinsic mode function (IMF) components. To solve this problem, Wu et al. [15] proposed an improved VMD method based on generalized demodulation technology to enhance the signal-to-noise ratio of decomposed components. This method was effectively applied to the identification of unbalanced faults and rotor crack faults. To screen IMF components, Liu and Tang [16] proposed an improved VMD method based on mutual information criteria. This method can accurately and reliably diagnose oil turbulence faults. Zhang et al. [17] proposed a method based on VMD-wavelet threshold and support vector machine (SVM) with particle swarm optimization (PSO). This method has an accuracy of 95% for unbalance fault diagnosis. The aforementioned VMD methods have important practical application value in rotor system fault identification. However, they have not solved the problem of relying on empirical knowledge to select penalty factor $\alpha$ and decomposition layer $V$. To ensure a more reasonable selection of parameters, Liu et al. [18] proposed a turbo-rotor fault identification method based on improved VMD. Within the method, the correlation coefficient criterion is applied to adaptively select the value of $V$. Wu et al. [19] proposed a failure feature analysis method with VMD and autoregressive spectrum. The instantaneous frequency mean method was used for screening the value of $V$. This method can effectively identify misalignment, unbalance, and loose faults. However, rather than optimizing the penalty factor $\alpha$ value, the authors based it on empirical values.

The fault diagnosis method based on feature extraction contains three main steps: data processing, feature extraction, and fault identification. The construction of the feature set is a key step that directly determines the accuracy of fault diagnosis [20]. When the rotor system of a rotating machine is exposed to a complex environment with strong background noise, the system often exhibits nonlinear and non-stationary characteristics. For such instances, the statistical feature parameter extraction based on signal processing can hardly accurately identify the fault feature. As a measure of complexity, the entropy index has also been widely applied in the field of fault pattern diagnosis. Azami et al. [21] proposed a refined composite multi-scale dispersion entropy (RCMDE). This is a scalar representing the complexity of the sequence which quantifies nonlinearity as well as non-stationarity of the signal. Zuo et al. [22] proposed an aeroengine rotor fault identification method based on the Fourier decomposition method and RCMDE. Within this method, multiple types of mixed faults can be effectively identified. However, only a single scale RCMDE value is selected as the eigenvalue. As such, the method is not suitable for practical engineering cases under different working conditions. Wang et al. [23] proposed a method based on complete ensemble empirical mode decomposition with adaptive noise and RCMDE. This method can accurately diagnose compressor faults.

However, vibration signals of the rotor system with different equipment, various environmental conditions, and diverse degrees of failure are different. Therefore, when the data distribution spaces of source and target domains are different, a transfer learning algorithm has to be used to improve the fault identification accuracy rate. Gopalan et al. [24] proposed a sampling geodesic flow (SGF) method based on Grassmann manifold. On this basis, Gong et al. [25] proposed the Geodesic Flow Kernel (GFK) method, which can adaptively select the optimal source domain. Zhang et al. [26] proposed a CNN network structure using a global mean pooling layer instead of a fully connected layer to realize transfer diagnosis under small samples. Lei et al. [27] summarized the development of domain adaptation methods in recent years, and pointed out that the future research direction is to extend it to engineering application scenarios. Considering actual engineering application conditions, transfer component analysis (TCA) is a very effective feature transfer method [28]. As such, it is often used to realize cross domain fault identification. Xiao et al. [29] proposed a method using improved denoising autoencoder and TCA. The proposed method has high fault diagnosis accuracy for vibration data in a noisy environment. Xu et al. [30] proposed a method based on multi-dimensional fault energy feature combined TCA and an optimized least squares support vector machine (LSSVM). This method
can effectively screen the fault information under different operating conditions. Chen et al. [31] proposed a transfer learning strategy based on an improved LSSVM model, which can achieve diagnostic tasks with a small amount of target domain data. However, there is still some room for improvement regarding recognition accuracy.

Adaptive boosting (AdaBoost) algorithm has been successfully applied to classification problems [32]. Multiple weak classifiers are combined to form a strong classifier through a weighted combination which can improve the performance of a single classifier. The combination of AdaBoost with machine learning methods has been effectively applied in the field of failure pattern recognition. Cao et al. [33] proposed a multi-classification algorithm based on AdaBoost-SVM. This method can be applied to monitor engine faults. However, kernel parameter $\sigma$ and penalty coefficient $C$ are still selected based on empirical values. To solve this problem, Hao and Jiao [34] proposed a wind turbine failure recognition method based on PSO-AdaBoost-SVM. PSO is applied to adaptively screen the values of $C$ and $\sigma$, thereby improving the improve fault recognition accuracy. However, as an improved version of SVM, LSSVM is more effective in classifying non-linear and high-dimensional data. Furthermore, LSSVM is characterized by a faster learning speed and is more suitable as a weak classifier of the AdaBoost algorithm.

In allusion to the problems of fault pattern recognition technology in industrial applications, the multi-source domain improvement fault diagnosis (MSDIFD) method is proposed. Considering that the vibration data of the rotor system comes from different equipment and working conditions in actual engineering, the label data under multiple equipment and working conditions are used as training data to enhance the generalization performance in multi-source fields. Then, the improved adaptive variational mode decomposition (IAVMD) method is applied to construct a joint feature screening index to screen sensitive components, reduce the noise interference of engineering data, and strengthen the extraction of fault information. Next, the optimal feature mapping between cross domain data is achieved based on weighted semi-supervised transfer component analysis (WSSTCA). In the end an ensemble failure mode classifier is built for online diagnosis on real-time data. The verification results show that this method has a good fault recognition effect and generalization. The contributions of this paper are as follows:

1. The MSDIFD method for rotor system typical faults automatic identification is proposed. It can be used for cross equipment, cross working condition, and cross domain rotor system diagnostic tasks.

2. The MSDIFD proposed in this paper realizes multidimensional improved fault diagnosis based on four aspects: source domain space, feature screening, feature transfer, and failure recognition. A domain generalization method is provided for cross domain diagnosis problems with incomplete target domain information.

3. The input of this method is the raw vibration waveform data of different equipment and different working conditions, and the output is failure recognition conclusions without relying on the empirical knowledge of external experts. The method has good cross domain transfer diagnosis performance and high identification accuracy.

The remaining sections of this paper are arranged as follows. The proposed method is introduced in Section II. The construction method for the typical fault automatic identification model of the rotor system is shown in Section III. A sufficient amount of multi-domain data of rotor system typical fault engineering cases (such as shaft misalignment failure, rotor unbalance failure, oil whirl failure, rubbing failure, and surge failure) are applied to validate the proposed method in Section IV. Finally, conclusions of this paper are described in Section V.

II. THEORETICAL BACKGROUND

A. MULTI-SOURCE DOMAIN FEATURE SPACE CONSTRUCTION

The research object of this paper is the rotor system in the actual engineering environment, and there is often no fault data of the running equipment. For the scenario where the fault information of the target domain is lacking in the model training stage, the commonality fault characteristics of the training data in the multi-source domain are mined to realize the pattern recognition of target domain samples, thereby improving the generalization of cross domain diagnosis.

![FIGURE 1. The multi-source domain feature transfer schematic diagram.](image_url)

At present, solving the problem of fault diagnosis under variable working conditions is the current research hotspot, but many scholars’ definitions of variable working conditions are not comprehensive enough. For example, merely several sets of bearing data with different damage degrees and different loads from the CWRU test bench are combined to construct a mixed variable working condition data set, and then divided into a training set and a test set according to a certain proportion to verify the model. This kind of method will face two main problems when it is applied in engineering:
(1) The data to be diagnosed is not just homologous data from the same equipment. For example, although they are all rotor systems, they are likely to come from centrifugal compressors and flue gas turbines which are two completely different equipment. In addition, working pressure, working speed, working temperature and other working conditions of the same type of equipment are very different. The difference of these real engineering cases cannot be simulated by the data of one test bench alone. (2) When diagnosing faults in actual cases, the working conditions of the data to be tested and the training data are very different, and they are often not from the same equipment. At this time, the model constructed by using a single equipment data working condition cannot extract the commonality features of the data from different equipment, which is easy to cause cross equipment and cross working condition data transfer failure, resulting in unsatisfactory diagnosis accuracy.

As shown in Fig. 1, the multi-source domain feature space in this paper includes n source domains from N different equipment: \( D_S = \{ D_S^1, D_S^2, \ldots, D_S^n \} \). In order to meet the practical scene of engineering application, the target domain also includes data of m target domains from M equipment: \( D_T = \{ D_T^1, D_T^2, \ldots, D_T^m \} \). Aiming at the above problems, a multi-source domain feature space is constructed based on n working conditions of N equipment, and after learning commonality knowledge, the features are transferred to cross equipment and cross working conditions data for diagnosis. Therefore, the test set in this paper is also the target domain data from m working conditions of M equipment which different from the training set. This method can solve the problem of the large gap between the target domain information and the source domain information to a certain extent, and improve the domain generalization ability for different target domain data.

**B. ADAPTIVE FEATURE EXTRACTION AND SCREENING**

Adaptive feature extraction and screening are performed using the IAVMD method shown in Fig. 2. The central theory of signal decomposition is that the decomposed mode components are all concentrated near the center frequency [35]. The transformed corresponding constrained variational model can be represented as:

\[
\min_{\{u_v\}, \{\omega_v\}} \sum_{v=1}^{V} \left\| \alpha \left( \delta (t) + \frac{j}{\pi t} \right) u_v (t) e^{-j\omega_v t} \right\|_2^2 \\
\text{s.t.} \sum_{v=1}^{V} u_v = y(t)
\]

where \( V \) is the decomposition layer, \( \{u_v\} \) is the \( V \)-th IMF components obtained by decomposition, \( \{\omega_v\} \) is the center frequency of each component, \( \delta \) is the Dirac distribution, \( y(t) \) is the raw signal, and \( t \) is the sampling time.

Lagrange multiplier \( \lambda \) and quadratic penalty factor \( \alpha \) are introduced to transform Eq. (1) into a non-constrained variational problem. As shown in Eq. (2), the Lagrange function expression is obtained as follows:

\[
L((u_v), \{\omega_v\}, \lambda) \\
= \alpha \sum_{v=1}^{V} \left\| \alpha \left( \delta (t) + \frac{j}{\pi t} \right) u_v (t) e^{-j\omega_v t} \right\|_2^2 \\
+ \left\| y(t) - \sum_{v=1}^{V} u_v (t) \right\|_2^2 + \left\{ \lambda, y(t) - \sum_{v=1}^{V} u_v (t) \right\}
\]

The alternating direction method of multipliers is used to update the process iteratively. Decomposition layer \( V \) and penalty factor \( \alpha \) have a vital impact on the decomposition effect, and cannot be selected based solely on empirical values. In addition, when practical engineering applications face noise interference, the selection of a single IMF component is sometimes too sensitive. For actual fault cases of cross equipment and cross working conditions, it is necessary to adaptively select sensitive IMF components containing valid fault information based on data. Therefore, it is necessary to screen and reconstruct the decomposed IMF components.

**FIGURE 2. The IAVMD method flowchart.**

The algorithm [36] with simple structure, strong robustness and fast convergence speed is required for engineering applications to adaptively optimize the values of \( V \) and \( \alpha \) according to the data, and then the \( \text{IK}_0 \) index is used to adaptively screen the sensitive IMF components. The IAVMD method can automatically output the feature reconstruction signal based on the raw vibration signal. This method maximizes the use of data feature information and minimizes the impact of empirical errors. In order to avoid the limitation of a single indicator, two indicators that have common characteristics for multi-source domain data of different equipment and working conditions are considered comprehensively. The correlation coefficient represents the degree of correlation between the decomposed IMF component and the raw vibration signal. The larger the correlation coefficient is; the more fault information the IMF component contains.
The energy ratio represents the energy proportion of the IMF component to the raw signal. When the rotor system fails, part of the fault shock will be converted into energy, so the energy ratio also contains abundant fault information. As shown in Fig. 2, an index combining energy ratio and the correlation coefficient is proposed to adaptively extract sensitive fault features. The specific steps are as follows:

1) The correlation coefficient \( r_v \) between the \( V \)-th IMF component \( f_v(t) \) and its raw signal \( s(t) \) is defined as:

\[
 r_v = \frac{\sum_{i=1}^{\infty} (s(t) - \bar{s}) (f_v(t) - \bar{f}_v)}{\sigma_s \sigma_v}
\]

where \( \bar{s} \) and \( \bar{f}_v \) are the average value of \( s(t) \) and \( f_v(t) \), respectively, while \( \sigma_s \) and \( \sigma_v \) are deviations of \( s(t) \) and \( f_v(t) \), respectively.

2) The energy ratio \( e_v \) of the \( v \)-th IMF is calculated as:

\[
 e_v = \frac{E_v}{E} = \frac{\int_{-\infty}^{+\infty} [f_v(t)]^2 \, dt}{\int_{-\infty}^{+\infty} |s(t)|^2 \, dt}
\]

where \( E_v \) is the energy occupied by the \( v \)-th IMF, and \( E \) is the total energy of the raw signal.

3) A sensitive IMF component evaluation index \( I_{RE} \) is proposed in the form of:

\[
 I_{RE}(v) = \frac{r_v}{\sum_{v=1}^{V} r_v} + \frac{e_v}{\sum_{v=1}^{V} e_v}
\]

After obtaining the reconstructed signal, the entropy values based on multi-scale analysis and temporal refinement are calculated as feature values [21]. Compared with the traditional time domain, frequency domain, and entropy features, RCMDE is less affected by occasional abnormal point fluctuations in equipment operating conditions, and is less sensitive to differences between different equipment and different operating conditions. The main factor affecting the value of RCMDE is the different fault types, so RCMDE is suitable as the commonality feature value for cross equipment and cross working conditions fault diagnosis. This method can extract the fault features of different signal scales, effectively reduce the loss of statistical information during signal processing, and reduced calculation deviation. The entropy value is calculated by the following steps.

1) As shown in Eq. (6), \( u(i), i = 1, 2, \cdots, L \) is the raw signal with data-length of \( L \). Scale factor \( \tau \)'s \( k \)-th sequence is given by:

\[
x^{(k)}_{j, \tau} = \frac{1}{\tau} \sum_{t=k \tau + (t-1) \tau}^{k \tau + j \tau - 1} u_t, \quad 1 \leq j \leq L, \quad 1 \leq k \leq \tau
\]

2) Feature value for each scale \( \tau \) can be calculated as follows:

\[
 E = -\sum_{\lambda=1}^{m} \ln(\overline{P}(\lambda_{u_0 u_1 \cdots u_{m-1}}))
\]

where \( \overline{P}(\lambda_{u_0 u_1 \cdots u_{m-1}}) \) is the probability of coarse-grained sequence, \( d \) is the time delay, and \( m \) is the embedding dimension.

C. WEIGHTED FEATURE TRANSFER

Feature transfer is a method that combines training data features with the current working condition signal [28]. For commonality features extracted from different equipment and working condition data, if they are directly used for classification and identification, it is difficult to obtain good classification performance in low-dimensional linear space due to the difference in data interval between different source and target domains. As a semi-supervised domain adaptation method, WSSTCA makes full use of the label information of the training data, the weighted kernel function is used to improve the transfer performance of the feature space and alleviate the negative transfer. This method takes into account the variability of various data cross equipment and cross working conditions, which can weaken the complexity requirements of the model to a certain extent. Maximum mean discrepancy is calculated as follows:

\[
dist(X_S, X_T) = \left\| \frac{1}{n_S} \sum_{i=1}^{n_S} \varphi(x_S) - \frac{1}{n_T} \sum_{i=1}^{n_T} \varphi(x_T) \right\|
\]

where \( \varphi \) is the mapping transformation, \( X_S \) is the source domain input, and \( X_T \) is the target domain input. As shown in Eq. (9), the distance between \( X_S \) and \( X_T \) can be converted as:

\[
dist(X_S, X_T) = tr(KL)
\]
Compared with unsupervised learning, WSSTCA method maximizes the dependence of the kernel function on the training data labels. A semi-supervised learning environment is built by encoding label information into learned components and preserving their local geometry. The manifold learning built by encoding label information into learned components is introduced to propagate label information from labeled data labels. A semi-supervised learning environment is maximized the dependence of the kernel function on the training data labels.

\[ H = E_{n_s + n_T} - [1/(n_s + n_T)]I \]

where \( H \) is the center matrix, \( K_{SS} \) is the source domain feature sample kernel matrix, \( E \) is the identity matrix, and \( I \) is the column vector of all ones.

As shown in Eq. (13), a constraint function is constructed to minimize the distance between \( x_a^s \) and \( x_b^s \) obtained by transforming the input samples \( x_a \), \( x_b \) through the feature mapping.

\[
\begin{align*}
    & \sum_{a,b} m_{ab} || x_a^s - x_b^s ||^2 = tr(J^T K \mathcal{L} K J) \\
    & m_{ab} = \exp(-d_{ab}^2/2\sigma^2), \quad \mathcal{L} = D - M
\end{align*}
\]

where \( \mathcal{L} \) is the Laplacian matrix, \( D \) is the diagonal matrix, \( M = [m_{ab}] \), and \( d_{ab} \) is the distance between \( x_a \) and \( x_b \) in the original input space.

In summary, combining Eq. (11), Eq. (12), and Eq. (13), the final objective function is:

\[
\begin{align*}
    & \min_w \ tr(J^T K \mathcal{L} K J) + \mu tr(J^T J) + \frac{\eta}{(n_s + n_T)^2} tr(J^T K \mathcal{L} K J) \\
    & s.t. \ J^T K H \bar{K}_{SS} \bar{K}_{J} = E, \quad \bar{K}_{SS} = \gamma K_{SS} + (1 - \gamma) E
\end{align*}
\]

where \( \mu \) is the regularization parameter, \( \eta \geq 0 \) and \( \gamma \geq 0 \) are tradeoff parameters.

In industrial applications, there are often data imbalances under different environments and working conditions. When a single kernel function is used, the transfer effect is not ideal. To solve this problem, the WSSTCA proposed in this paper by a variety of kernel functions to perform convex combination. A weighted kernel function is constructed to map the commonality sample features extracted from the multi-source domain feature space to a high-dimensional Hilbert space, and achieve maximum feature alignment of source and target domain in the high-dimensional space. The differences between cross domain data across equipment and working conditions are minimized, achieving optimal feature knowledge transfer and improving the accuracy of transfer diagnosis.

The construction method of the weighted kernel function is:

\[ K(x_p, x_q) = \sum_{r=1}^{R} k_r K_r(x_p, x_q) \]

where \( R \) is the number of kernel functions, and \( k_r \) is the kernel function weight.

Overall Considering Mercer’s theorem, the distribution characteristics of actual engineering data and the needs of engineering applications, a Gaussian radial basis kernel (RBF) function with few parameters and multiple decision boundaries and a Sigmoid kernel function with good generalization performance for unknown samples are used to construct a weighted kernel function.

\[ K(x_p, x_q) = kK_{RBF} + (1 - k)K_{SIG} \]

where \( k \) is the weighting factor and the algorithm [36] is applied to adaptively determine the \( k \) value based on the feature value calculated from the real-time data. \( K_{RBF} \) is the Gaussian RBF, and \( K_{SIG} \) is the Sigmoid kernel function.

Finally, the optimal mapping kernel matrix \( J \) can be obtained by WSSTCA method. The matrix \( J \) contains the transferred features that minimize the data distribution distance by mapping the source domain features and the target domain features. At the same time, this matrix propagates the label information in the source domain knowledge to the target domain based on maximizing the label correlation. Therefore, the matrix \( J \) can be directly input into the classifier to identify the fault classification corresponding to the target domain data.

**D. ENSEMBLE FEATURE CLASSIFICATION RECOGNITION**

In the ensemble feature classification algorithm, multiple weak classifiers are constructed, the parameters of each weak classifier are determined adaptively according to the data. Since the research in this paper is aimed at cross working condition and cross equipment typical fault diagnosis, the training data is multi-source domain data from different equipment and working conditions, and the working conditions of the test data are also quite different from the training data. Therefore, when constructing a classifier, it is necessary to continuously adjust the weight of the weak classifier which accurately identify the training samples based on the commonality features extracted from the source domain data, and continuously generate a classifier with slightly better performance.

The weak classifier is continuously updated iteratively based on the training data, and finally weighted into a strong classifier. Based on the self-similarity and difference within the data feature space, this method can obtain a more comprehensive failure mode classifier. A specific flowchart of the ensemble feature classification algorithm proposed in this paper is shown in Fig. 3.

The classifier solves the linearly inseparable problem by mapping the extracted features to a high-dimensional space for expression [37]. Based on the good processing ability of high-dimensional linear inseparable data, N LSSVMs are constructed and given the same initial weight as the underlying weak classifier. For input training samples of weak classifiers\([(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)]\), the optimization
problem can be described as:

\[
\min J(\theta, \beta) = \frac{1}{2} \| \theta \|^2 + C \sum_{i=1}^{N} \beta_i^2 \\
\text{s.t. } y_i[\theta^T \varphi(x_i) + \alpha] = 1 - \theta_i
\]  \tag{17}

where \( \theta \) is the weight vector, \( \beta \) is the fitting error, \( C \) is the penalty coefficient, \( y_i \in \{-1, 1\} \) is the mapped value of the training sample, \( x_i \) is the input value of the training sample, and \( \alpha \) is the offset vector.

The initial parameter values of the sparrow search algorithm is employed to optimize the parameter selection based on empirical values and balance the performance increases and vice-versa. To solve the problem with an increase in the value of \( \sigma \), \( C \) remains unchanged, classification performance worsens.

The main purpose of this paper is to propose the MSDIFD and cross domain fault diagnosis tasks. The corresponding mathematical model is defined as follows:

\[
X_{i,j}^{c+1} = \begin{cases} 
X_{best}^{c} + \xi & \text{if } f_p > f_b \\
X_{i,j}^{c} + W \left[ X_{i,j}^{c} - X_{worst}^{c} \right] & \text{if } f_p = f_b 
\end{cases}
\]  \tag{21}

where \( X_{best} \) and \( X_{worst} \) represent the current best and worst global optimal location respectively, \( \xi \) is the step size, \( c \) is the current iteration number, \( f_p \) is the current fitness value, \( W \) is a random number, \( f_b \) and \( f_w \) are the current best and worst fitness values, respectively.

Finally, the fitness value is calculated and the entire population position is updated to determine whether the iteration termination conditions are met, and the optimal parameter values \( C \) and \( \sigma \) are obtained. The optimized parameter values obtained adaptively are substituted into the constructed weak classifiers, and the weights of each weak classifier are iteratively updated based on the recognition results of the training samples. Ultimately, a strong classifier is output by weighted combination. This classifier is used to identify commonality features after transfer and mapping, so as to realize cross equipment, cross working conditions, and cross domain fault diagnosis tasks.

### III. PROPOSED FAULT DIAGNOSIS METHOD

The main purpose of this paper is to propose the MSDIFD method which can be applied to online state monitoring and automatic fault identification for the rotor system of rotating machinery.

In Fig. 4, the green part is the adaptive feature extraction process in the offline training mode. The input of the mode is five types of fault training data sets of five typical equipment pieces. The output is the classification corresponding to feature values of the five types failure modes. The red dashed box represents the online diagnosis mode of the method, the blue part in the box is the dynamic fault recognition knowledge base, and the yellow area is the weighted feature.
transfer. Its input is the real-time raw vibration waveform data and output is the fault mode identification conclusion.

A. VIBRATION SIGNAL DECOMPOSITION AND RECONSTRUCTION

The vibration signal of the rotor system is often characterized by weak energy, non-stationarity, non-linearity, and noise interference. IAVMD, as a non-recursive method, can effectively deal with the non-linearity and non-stationarity of the signal. Furthermore, this algorithm overcomes the problem of selecting the $V$ and $\alpha$ according to empirical values, and parameter values can be adaptively determined for different signals.

It can be seen from Section II that the method for screening IMF components that contain valid fault information is of great significance for constructing an accurate fault diagnosis model. The sensitive fault information is extracted based on the joint analysis index $I_{RE}$ constructed in this paper.

The $I_{RE}$ values are sorted from highest to lowest. Three IMFs with the largest $I_{RE}$ values are selected as sensitive components for reconstruction. The remaining components are regarded as noise components and removed. This method can retain effective components containing failure features, denoise the raw waveform signal, and increase the accuracy of fault identification.

B. FEATURE VALUE CALCULATION AND TRANSFER

The feature value of the reconstructed signal is associated with the following parameters [21]. Embedding dimension $m$ affects the sensitivity to signal fluctuations, while its value in this paper is three. Category $c$ has a significant influence on the difference in capturing amplitude. It is set to six in this paper. The time delay $d$ is one. If the value of $d$ is overly high, the signal will be lost due to untimely transmission. In addition, as shown in Eq. (6), when $e^{\alpha m} < L/\tau$ is satisfied, the calculation result is reliable. In order to meet the application requirements of engineering data for different equipment and working conditions, and to better characterize the feature information of different fault types, set $\tau = 15$.

Relevant researches on fault diagnosis transfer include three types: data-based transfer, model and parameter-based transfer, and feature-based transfer [39]. All three transfer methods have solved the problem of failure recognition to some extent under different equipment and working conditions. However, not all of the aforementioned methods are effective in industrial applications. Once the data distributions are significantly altered under approximate working conditions, data transfer may cause relatively high errors in fault diagnosis. The transfer method of model and parameters is too dependent on the model itself. As a feature transfer method that combines training data with current working condition data, WSSTCA simultaneously considers data similarity under different operating conditions and applies the fault diagnosis models to engineering applications. First, the historically known label signal and the current real-time working condition signal are simultaneously spatially transformed. Then, the feature values of both training data and of ones calculated in real-time are transformed and combined into a new feature vector matrix. Due to lower differences, the dependence of pattern recognition on the model is weakened.

In order to retain the complete fault feature information, avoid information loss caused by data dimensionality reduction, and improve the fault diagnosis ability of cross domain data, the dimension of the feature matrix after WSSTCA transformed is set to five.

C. KNOWLEDGE BASE CONSTRUCTION AND FAILURE RECOGNITION

Five fault source categories are employed in this paper. The vibration signal is trained according to the aforementioned steps, so as to obtain the knowledge base described in Fig. 4 composed of the feature values of five fault modes. The categories in the knowledge base for shaft misalignment, rotor unbalance, oil whirl, rubbing, and surge faults correspond to denotes as 1, 2, 3, 4, and 5, respectively. By comprehensively considering abundant rotor system fault case, engineering application experience, and
feasibility of cross domain diagnosis coupled with strong noise in different equipment and working conditions, a failure mode recognition rule is determined in this paper. If the fault identification accuracy rate of the real-time test data reaches 80% or above, the current unit fault mode can be determined. If the fault recognition accuracy rate is below 80%, the fault mechanism should be analyzed. The data of the new failure mode determined by the mechanism analysis can be added to the knowledge base based on above steps to improve the generalization ability of the model, so that a constantly improving and enriching dynamic fault identification knowledge base is constructed.

The ensemble feature recognition classifier is used to achieve multi-class recognition of typical faults in the rotor system. All training samples are set to the same initial weight with the sum of one. A total of 100 weak classifiers are set in this paper, and the initial weight of each classifier is 0.01. To ensure that the next iteration focuses on learning the wrong sample, the weight of the correct samples is reduced and the weight of the wrong samples is increased. As shown in Fig. 3, if the identification error rate $\varepsilon_i < 0.5$ of the weak classifier, the iteration is completed for subsequent combination processing, and if the error rate $\varepsilon_i \geq 0.5$, the iterative classification process is continued until the conditions are met. Then, the weights of all weak classifiers are calculated and the sample weights are updated accordingly, and each weak classifier is combined into a strong classifier based on the normalized weights.

The “one versus one” principle is used to extend the single weak classifier to multiple classifiers. In other words, two types of training samples are selected to construct a classifier. For five classification problems in this paper, ten classifiers need to be established. If $\sigma$ is too small, over-fitting is likely to occur. On the other hand, if $\sigma$ is too large, the classifier is not strong enough to meet the classification conditions [40]. Therefore, in this paper, the ensemble feature classification algorithm is employed to adaptively determine the values of penalty coefficient $C$ and kernel parameter $\sigma$. Moreover, it can continuously generate slightly better performing classifiers during iterations. Finally, a strong classifier is generated to achieve failure mode recognition.

### IV. ENGINEERING VERIFICATION AND METHOD COMPARISON

Obtaining the state monitoring data of the entire performance degradation process of the rotating machinery rotor system from “normal operation to failure shutdown” is of great value for training and testing the fault recognition model. With the help of Shenzhen SBW Monitoring and Control Tech. Co., Ltd, the research team obtained rotor systems typical failure label data of centrifugal compressors, steam turbines, and flue gas turbines. These components include five types of typical fault case data: shaft misalignment, rotor unbalance, oil whirl, rubbing, and surge.

### A. INTRODUCTION TO THE TRAINING DATASET

Five types of typical failure data for practical failure cases from five different petrochemical companies are shown in Table 1. The Eddy current displacement sensor is used to collect raw vibration data. Since the rotation speed of the rotor system in industrial applications is not constant, the equal-angle sampling method is adopted. A total of 32 data points are collected per revolution for each data file, and the sampling points is $32 \times 32 = 1024$. According to the Shannon-Nyquist sampling theorem [41], the sampling frequency of a signal is as follows:

$$f_s = \frac{P \times S}{60}$$

where $P$ is the number of points collected per revolution, and $S$ is the real-time speed.

As shown in Fig. 5, the XJ-090-014 unit consists of steam turbine (in the left) and centrifugal compressor (in the right). The shaft misalignment faults occurred during the unit operation, and fault characteristic of vibration measurement points is particularly obvious at the coupling end of centrifugal compressor.

As shown in Fig. 6, the YX-397-033 unit is composed of a steam turbine (in the left), a high-pressure cylinder (in the middle) of centrifugal compressor, and a low-pressure cylinder (in the right). During the operation of the centrifugal compressor unit, abnormal vibration of measuring points was detected, and rotor unbalance fault occurred.

As shown in Fig. 7, the XA-443-018 unit is made up of steam turbine (in the left) and centrifugal compressor (in the right). The steam turbine of unit has an oil whirl failure.

As shown in Fig. 8, YL-131-015 is a steam turbine unit. There are vibration fluctuations at the high-pressure non-coupling end of steam turbine, and the rubbing fault has been occurred.

As shown in Fig. 9, measurement points’ vibration amplitude of the centrifugal compressor unit SX-002-010 during operation has significantly increased close to the alarm threshold, and there is a tendency to continue deteriorating. This unit has the characteristics of surge fault.

For each type of fault, such as shaft misalignment, rotor unbalance, oil whirl, rubbing, and surge, 150 sets of data files...
are selected for each. Thus, a total of 750 data file sets from the training data set are obtained. The vibration waveform and frequency spectrum of the selected typical fault data are shown in Fig. 10 and Fig. 11, respectively.

As shown in Fig. 11, the dominant frequency is twice as high as the fundamental frequency (FF) in frequency spectrum of shaft misalignment fault (FF$_1$ = 68.3 Hz). The Fundamental frequency of rotor unbalance fault frequency spectrum is the dominant frequency (FF$_2$ = 128.3 Hz). When the oil whirl failure occurs, the whirl frequency is the dominant frequency (FF$_3$ = 176.7 Hz). In addition, the accompanying frequency is the same as the fundamental frequency. In the rubbing fault frequency spectrum, the dominant frequency is four times higher than the fundamental frequency (FF$_4$ = 50 Hz). The accompanying frequency is the same as the fundamental frequency. Lastly, it should be mentioned that frequencies twice or three times as high are also present. When a surge fault occurs in the rotor system, obvious low-frequency signal components appear generally within 1-30 Hz (FF$_5$ = 141.7 Hz). In conclusion, five types of representative fault data that constitute the training data set have typical and distinguishing fault characteristics that can be applied to model training and testing.

B. FAULT IDENTIFICATION MODEL

The IAVMD algorithm is applied to process the raw signal of the training data set. The raw waveform signal is adaptively decomposed into six layers of IMF components. The I$_{RE}$ value of six-layer IMF components is obtained after the vibration signal is decomposed and calculated.

As shown in Fig. 12, I$_{RE}$ values of the first three IMF components are significantly higher than the other components. These components are sensitive and contain a significant amount of fault information. Therefore, three IMFs with the largest I$_{RE}$ values are selected for reconstruction. The remaining IMF components are not highly correlated with the raw signal, have a small proportion of energy, and contain less feature information. Insensitive feature components are thereby filtered out.

As shown in Fig. 13, the feature mean value of 150 sets of data files for each of five typical faults is calculated.
FIGURE 11. Frequency spectrum of the training data.

FIGURE 12. IMF component $I_{Eg}$ value.

On the first five scales, the entropy relationship of rotor system vibration signal in five fault states is $E$ (Surge) > $E$ (Oil whirl) > $E$ (Rubbing) > $E$ (Shaft misalignment) > $E$ (Rotor unbalance). The higher energy value, the lower self-similarity of the signal. If the feature value of too many scales is selected, it will cause information redundancy, and the calculation time will be too long, which is not conducive to industrial applications. On the contrary, too few scales cannot contain complete fault information, which will affect the recognition accuracy. Five different faults can be distinguished on the first five scales, and the entropy value of each fault state has begun to decrease after the sixth scale, indicating that the first five scales already contain sufficient fault information. Consequently, feature values of the first five scales are screened to construct the feature vector matrix in this paper.

As shown in Table 2, the data for five typical engineering failure cases different from the training set data are selected as the model test set. Five typical engineering failure modes include shaft misalignment, rotor unbalance, oil whirl, rubbing, and surge. Each type of engineering failure case data contains 100 data files, each 1024 in length. The test set contains 500 sets of data files.

The vibration waveforms and frequency spectrums of 100 data files of the same failure mode have certain similarity. Therefore, as shown in Fig. 14 and Fig. 15, only the vibration waveform and frequency spectrum of a set of data for each typical fault type in the test data are drawn.

Based on the aforementioned five types of fault data, the ensemble feature classification algorithm is applied to adaptively determine the optimal value of kernel parameter $\sigma$ and penalty coefficient $C$. The results are shown in Fig. 16. In addition, the optimal adaptive value of the weighting factor $k$ is 0.55.

As shown in Fig. 16, optimal parameter values are very different for five different types of fault data. If parameters are set based on empirical knowledge, large errors will occur and affect the fault identification accuracy. Consequently, it is necessary and effective to employ the ensemble feature classification algorithm to adaptively determine the global optimal value of penalty coefficient $C$ and kernel parameter $\sigma$. The recognition accuracy of the test set is shown in Fig. 17.

As shown in Fig. 17, the model has an accuracy of 98% for identifying shaft misalignment fault. Moreover, the fault
modes represented by two labeled data files are incorrectly identified as static and dynamic rubbing. The identification accuracy of the model is 100% for the rotor unbalance fault data. The accuracy rate is 99% for identifying oil whirl fault and surge fault data. For both of these failure modes, each has a labeled data file that was incorrectly identified as surge and oil whirl, respectively. Three labeled data files of rubbing fault were incorrectly identified as rotor unbalance and oil whirl faults. Therefore, the recognition accuracy is 97%. The model verification results represent that this typical failure pattern recognition model can accurately diagnose the failure data of the rotor system under different equipment and various working conditions. Furthermore, ablation experiments are designed to verify the effectiveness and necessity of the proposed method’s improvement in three parts: feature extraction, feature transfer, and fault identification. In this paper, IAVMD method is used to achieve adaptive feature extraction, WSSTCA method is applied to transfer the source domain information to the target domain, and fault diagnosis is completed based on the ensemble feature classification algorithm. Therefore, in ablation experiments, each individual optimization strategy...
and their combinations were validated separately. A total of 500 sets data of five fault types are used to conduct experiments, and the verification results are shown in Table 3.

As shown in Table 3, when no improvement method is used, the average fault diagnosis accuracy is the lowest, only 70.2%. When any optimization method is applied alone, the recognition accuracy will be improved to a certain extent. The effect of applying the weighted feature transfer method alone is relatively the best, with an average recognition accuracy rate of 82%.

There are no cases of negative optimization and mutual interference.

C. COMPARISON OF METHODS

1) COMPARISON OF FEATURE EXTRACTION METHODS
Feature extraction and screening are very critical steps in fault diagnosis, which directly determine whether the fault features in the raw data can be fully and effectively represented, thereby affecting the accuracy of fault identification. In order to demonstrate the effectiveness of the IAVMD method, based on the above selected 500 sets of test data files, four different feature extraction/dimensionality reduction methods are applied to compare the fault diagnosis accuracy while keeping the rest of the method same. Comparison method 1: The classical VMD method [35] is used to decompose the vibration signal, the feature vector is constructed and reduced the dimension with the principal component analysis (PCA) method. Comparison method 2: The PF components obtained by local mean decomposition (LMD) are reconstructed according to the correlation coefficient criterion [42]. Then the multipoint optimal minimum entropy deconvolution adjusted (MOMEDA) method is used to denote the reconstructed signal and extract the fault features. Comparison method 3: The intrinsic scale components decomposed by the local characteristic-scale decomposition (LCD) with a large correlation coefficient are selected for reconstruction, and then the maximum correlated kurtosis deconvolution (MCKD) is used to enhance the fault feature representation of the reconstructed signal [43]. Comparison Method 4: The fault feature extraction method based on ensemble empirical mode decomposition (EEMD) and independent component analysis (ICA) [44].

As shown in Table 4, when IAVMD is applied as the feature extraction and screening method, the fault identification accuracy rate on the test dataset is the highest. The results indicate that the IAVMD method can accurately separate the fault information in the practical engineering vibration signal, and effectively extract the fault characteristics of the rotor system. In the industrial application environment, it is more superior than the other four methods.

2) COMPARISON OF CLASSIFICATION RECOGNITION METHODS
To further reflect the superiority of the proposed method, five different classification recognition methods are selected for comparison. The rest parts of the comparison method are consistent with the MSDIFD method, except for the following classification recognition algorithm. A fault identification method based on the stacked denoising autoencoder [11] and TCA (defined as Method 1). A fault identification method including AdaBoost-SVM [37], and SSTCA [28] (defined as Method 2). A fault identification method containing GFK [25] (defined as Method 3). A fault identification method involving global average pooling CNN [26] (defined as Method 4). A fault identification method with TrAdaBoost [27] (defined as Method 5).

In Method 1, stacked denoising autoencoder combined with TCA is employed for deep transfer learning on the extracted RCMDE feature values to achieve fault recognition. In Method 2, AdaBoost-SVM optimized with PSO is applied to classify and recognize RCMDE feature values transformed by the SSTCA. In Method 3, the multi-source domain transfer learning GFK method with LSSVM as the baseline classifier combined with RCMDE for fault identification. In Method 4, a deep convolutional neural network based on a global mean pooling layer is used to extract the features of fault data to realize pattern recognition. In Method 5, the single-source domain transfer learning method TrAdaBoost is combined with the LSSVM classifier for failure pattern recognition.

To obtain the best performance, hyperparameters for each method are combined with the above references adjusted with cross-validation based on the training data. The test results are shown in Table 5. Based on these five methods and selected test data sets, the fault recognition accuracy of methods is mutually compared. The summary fault identification results of six different methods for the selected five types of typical failure data sets are shown in Table 6.

The average fault recognition accuracy of the six methods is compared horizontally, and Method 1 has the worst recognition effect on the test data set. The recognition accuracy rates of shaft misalignment, rotor unbalance, oil whirl, rubbing, and surge faults are 90%, 82%, 80%, 73%, and 69%, respectively. The average fault identification accuracy is 78.8%. The longitudinal comparison of the recognition accuracy of same faults by different methods shows that the five comparison methods have relatively high recognition accuracy for shaft misalignment and rotor
TABLE 1. Training data description.

| Label | Measuring point code | Fault state       | Sensor location                                      | Speed (r/min) | Sampling points | Data files number |
|-------|----------------------|-------------------|------------------------------------------------------|---------------|-----------------|-------------------|
| 1     | XJ-090-014           | Shaft misalignment | X channel at coupling end of centrifugal compressor  | 4100          | 1024            | 150               |
| 2     | YX-397-033           | Rotor unbalance   | X channel at coupling end of centrifugal compressor  | 7700          | 1024            | 150               |
| 3     | XA-443-018           | Oil whirl         | X channel at coupling end of steam turbine           | 10600         | 1024            | 150               |
| 4     | YL-131-015           | Rubbing           | X channel at non-coupling end of steam turbine       | 3000          | 1024            | 150               |
| 5     | SX-002-010           | Surge             | V1124 of centrifugal compressor                      | 8500          | 1024            | 150               |

TABLE 2. Test data description.

| Label | Measuring point code | Fault state       | Sensor location                                      | Speed (r/min) | Sampling points | Data files number |
|-------|----------------------|-------------------|------------------------------------------------------|---------------|-----------------|-------------------|
| 1     | ZK-933-002           | Shaft misalignment | Y channel at coupling end of centrifugal compressor  | 4500          | 1024            | 100               |
| 2     | ZT-026-062           | Rotor unbalance   | X channel at non-coupling end of centrifugal compressor | 3000          | 1024            | 100               |
| 3     | YT-620-017           | Oil whirl         | V11572B of steam turbine                              | 6100          | 1024            | 100               |
| 4     | NM-181-066           | Rubbing           | X channel at non-coupling end of steam turbine       | 4500          | 1024            | 100               |
| 5     | AK-093-015           | Surge             | Y channel at non-coupling end of centrifugal compressor | 11000         | 1024            | 100               |

TABLE 3. Test results of ablation experiments.

| Adaptive feature extraction | Weighted feature transfer | Ensemble feature classification | Recognition accuracy of five fault types (%) | Average (%) |
|-----------------------------|---------------------------|---------------------------------|---------------------------------------------|-------------|
| NO                          | NO                        | NO                              | Shaft misalignment: 74.0 Rotor unbalance: 71.0 Oil whirl: 69.0 Rubbing: 75.0 Surge: 62.0 | 70.2        |
| YES                         | NO                        | NO                              | Shaft misalignment: 78.0 Rotor unbalance: 75.0 Oil whirl: 69.0 Rubbing: 79.0 Surge: 68.0 | 73.8        |
| NO                          | YES                       | NO                              | Shaft misalignment: 88.0 Rotor unbalance: 83.0 Oil whirl: 79.0 Rubbing: 86.0 Surge: 74.0 | 82.0        |
| YES                         | NO                        | YES                             | Shaft misalignment: 84.0 Rotor unbalance: 80.0 Oil whirl: 77.0 Rubbing: 82.0 Surge: 70.0 | 78.6        |
| NO                          | YES                       | YES                             | Shaft misalignment: 92.0 Rotor unbalance: 90.0 Oil whirl: 87.0 Rubbing: 89.0 Surge: 85.0 | 88.6        |
| YES                         | NO                        | YES                             | Shaft misalignment: 90.0 Rotor unbalance: 88.0 Oil whirl: 85.0 Rubbing: 88.0 Surge: 85.0 | 87.2        |
| YES                         | YES                       | YES                             | Shaft misalignment: 98.0 Rotor unbalance: 100.0 Oil whirl: 99.0 Rubbing: 97.0 Surge: 99.0 | 98.6        |

TABLE 4. Test results of five feature extraction methods.

| Comparison method | Five types of fault recognition accuracy (%) | Average (%) |
|-------------------|---------------------------------------------|-------------|
| VMD-PCA           | Shaft misalignment: 96.0 Rotor unbalance: 93.0 Oil whirl: 92.0 Rubbing: 93.0 Surge: 91.0 | 93.0        |
| LMD-MOMEDA        | Shaft misalignment: 95.0 Rotor unbalance: 90.0 Oil whirl: 91.0 Rubbing: 89.0 Surge: 87.0 | 90.4        |
| LCD-MCKD          | Shaft misalignment: 89.0 Rotor unbalance: 90.0 Oil whirl: 85.0 Rubbing: 88.0 Surge: 87.0 | 87.8        |
| EEMD-JCA          | Shaft misalignment: 90.0 Rotor unbalance: 87.0 Oil whirl: 89.0 Rubbing: 91.0 Surge: 92.0 | 89.8        |
| LAVMD             | Shaft misalignment: 98.0 Rotor unbalance: 100.0 Oil whirl: 99.0 Rubbing: 97.0 Surge: 99.0 | 98.6        |

TABLE 5. Hyperparameter settings for each method.

| Method | Hyperparameter |
|--------|----------------|
| Method 1 | network layers=4, activation function=Sigmoid, noise reduction ratio=0.5, self-decaying learning rate=2 |
| Method 2 | weak classifiers = 100, penalty coefficient =10, regularization parameter=1, tradeoff parameter=0.5 |
| Method 3 | subspace dimensionality =10, scale parameter=0.5 |
| Method 4 | convolutional layer: convolutional kernels=4, kernel size=15, stride=1, pooling layer: size=2, stride=2 |
| Method 5 | iterations number=200, initial weight = 0.005, error factor=0.5 |

unbalance faults, and low recognition accuracy for oil whirl and surge faults. This indicates that these methods cannot accurately extract the fault characteristics with relatively complex failure mechanisms, and cannot learn the real commonality diagnostic knowledge from the source domain data, so they are unable to accurately identify the target domain fault data. However, the MSDIFD method proposed in this paper is very effective in identifying five types of faults, with an average accuracy of 98.6%. The results indicate that the fault recognition effect of the MSDIFD method is significantly better than the other five methods, and demonstrate that the method has good generalization for...
TABLE 6. The recognition accuracy of six methods for 5 types fault cases.

| Comparison method | Shaft misalignment | Rotor unbalance | Oil whirl | Rubbing | Surge | Average (%) |
|-------------------|--------------------|-----------------|-----------|---------|-------|-------------|
| Method 1          | 90.0               | 82.0            | 80.0      | 73.0    | 69.0  | 78.8        |
| Method 2          | 97.0               | 95.0            | 81.0      | 93.0    | 88.0  | 90.8        |
| Method 3          | 88.0               | 81.0            | 87.0      | 84.0    | 82.0  | 84.4        |
| Method 4          | 89.0               | 80.0            | 75.0      | 81.0    | 73.0  | 79.6        |
| Method 5          | 87.0               | 81.0            | 79.0      | 83.0    | 77.0  | 81.4        |
| MSDIFD method     | 98.0               | 100.0           | 99.0      | 97.0    | 99.0  | 98.6        |

failure mode recognition of the rotor system under different equipment and various operating conditions.

3) EFFECTS COMPARISON OF FAULT DIAGNOSIS METHODS

In order to further verify the influence of the increase of sample number on the fault identification accuracy of the proposed method in the industrial big data environment, another 40 sets of rotating machinery engineering failure case data are used as test data. Due to the different frequency of failures and different complex data acquisition, the number of data files obtained for each type of failure is also different. For example, centrifugal compressors are equipped with anti-surge systems. Consequently, their frequency of surge failures is low, and fewer failure data exist for each surge occurrence.

In the test data set, 1-10 data files are shaft misalignment fault data, 11-23 data files are rotor unbalance fault data, 24-32 data files are oil whirl fault data, 33-37 data files are rubbing fault data, 38-40 data files are surge fault data, and the data length of each file is 100. A total of 4000 sets of data files come from 15 different centrifugal compressors, flue gas turbines and other equipment. Their operating conditions such as measuring point information, speed, and load are different from the training data, and also different from the 500 sets of test data used above. The method proposed in this paper and the aforementioned five comparison methods are used to validate typical fault data across equipment and working conditions.

As shown in Fig. 18, for the recognition results of the MSDIFD method, only four sets of data have a fault recognition accuracy rate of less than 80% (one set each for shaft misalignment, rotor unbalance, oil whirl, and surge faults). Moreover, the recognition accuracy of the other five methods on these four data files is lower than the proposed method. This indicates that the four data files cannot accurately align the target domain and source domain feature spaces due to the excessive noise interference of their vibration signals, and the fault feature information is submerged or lost. Such instances are considered as fault data that cannot be accurately identified. However, the recognition accuracy is quite different based on the remaining 36 data files, but the recognition effect of the proposed method is still better than the other five methods. The above results show that for the “bad data” which lost fault information, the MSDIFD method can enhance the expression of fault features and improve the recognition accuracy compared with other methods. For typical fault data, commonality diagnostic knowledge can be mined to realize cross domain transfer diagnosis, and the fault identification accuracy will not drop significantly with the increase of samples in the target domain. From the perspective of overall failure recognition accuracy and single-set data recognition accuracy, the method proposed in this paper has a superior fault diagnosis effect.

As shown in Table 7, the average recognition accuracy of Method 1, Method 2, Method 3, Method 4, and Method 5 for 40 sets of typical fault data is 76.7%, 86.1%, 79.8%, 76.4%, and 78.6%. The standard deviations of the recognition accuracy for five methods are 10.4%, 10.3%, 9.3%, 9.8% and 9.0%, respectively. The aforementioned five methods have been validated by a relatively large amount of data. Their recognition accuracies are low while the standard deviation of test results are relatively high. It indicates that these methods are unstable and has significant differences in the fault recognition effect of the rotor system vibration data under different operating conditions. Consequently, it is difficult to achieve industrial applications when accompanied by strong noise background.
TABLE 7. The identification results of six methods for 40 groups fault cases.

| Comparison method | Recognition result (%) | Average accuracy | Standard deviation |
|-------------------|------------------------|-----------------|-------------------|
| Method 1          | 76.7                   | 76.7            | 10.4              |
| Method 2          | 86.1                   | 86.1            | 10.3              |
| Method 3          | 79.8                   | 79.8            | 9.3               |
| Method 4          | 76.4                   | 76.4            | 9.8               |
| Method 5          | 78.6                   | 78.6            | 9.0               |
| MSDIFD method     | 91.5                   | 91.5            | 8.8               |

The MSDIFD method has the highest average accuracy of failure recognition, reaching 91.5%. At the same time, the standard deviation is the smallest at 8.8%. The results indicate that the method provides a feasible way for typical fault pattern recognition of the rotor system under different equipment and operating conditions in the industrial environment. The MSDIFD method proposed in this paper has been validated by a sufficient amount of practical engineering fault cases, which is characterized by a good generalization and fault identification effects.

V. CONCLUSION

(1) The MSDIFD method for rotor system fault identification was proposed in this paper. It is progressively optimized from four aspects: multi-source domain feature space construction, adaptive feature extraction, weighted feature transfer, and ensemble feature classification and identification, which meets the needs of fault identification under different equipment or operating environments in practical engineering applications. This method can learn more comprehensive and generalized commonality fault information as much as possible, highlight fault feature representation, and realize cross equipment, cross working condition and cross domain fault diagnosis of rotor system.

(2) The method proposed in this paper has excellent fault identification accuracy for shaft misalignment, rotor unbalance, oil whirl, rubbing, and surge fault. The average recognition accuracy rate of MSDIFD method for 40 sets of rotating machinery rotor system fault data is 91.5% with the corresponding standard deviation of 8.8%. The results demonstrate that the proposed method is relatively stable for rotor system fault data in different industrial application environments, and has good domain generalization ability for different target domain data. The advantages of this method are proven by mutually comparing different fault recognition methods.

(3) Compared with the MSDIFD method, the average recognition accuracy of five classical fault identification methods for 40 sets of data files is 76.7%, 86.1%, 79.8%, 76.4%, and 78.6%, respectively. The corresponding standard deviations are 10.4%, 10.3%, 9.3%, 9.8% and 9.0%, respectively. Verification results indicate that the traditional fault identification methods have certain limitations in the multi-target domain fault diagnosis of different equipment and working conditions. The MSDIFD method can extract commonality features of complex fault mechanisms, and improve the accuracy of fault identification.

REFERENCES

[1] A. G. Nath, A. Sharma, S. S. Udmale, and S. K. Singh, “An early classification approach for improving structural rotor fault diagnosis,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–13, 2021, doi: 10.1109/TIM.2020.3043959.
[2] C. L. Li, M. Yun, Z. Rui, and W. Wei, “Rotor system fault diagnosis based on semi-supervised spectrum kernel clustering,” J. Vib. Meas. Diagnosis, vol. 36, no. 3, pp. 562–567, 2016.
[3] L. Xiang, Z. Deng, A. Hu, and X. Gao, “Multi-fault coupling study of a rotor system in experimental and numerical analyses,” Nonlinear Dyn., vol. 97, no. 4, pp. 2607–2625, Sep. 2019.
[4] V. R. Patil and P. V. Jadhav, “Dynamic response analysis of unbalanced rotor-bearing system with internal radial clearance,” SN Appl. Sci., vol. 2, no. 11, p. 1826, Nov. 2020, doi: 10.1007/s42452-020-03688-y.
[5] N. H. Chandra and A. S. Sekhar, “Fault detection in rotor bearing systems using time frequency techniques,” Mech. Syst. Signal Process., vol. 72–73, pp. 105–133, 2016.
[6] H. Shao, M. Xia, G. Han, Y. Zhang, and J. Wan, “Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer convolutional neural network and thermal images,” IEEE Trans. Ind. Informat., vol. 17, no. 5, pp. 3488–3496, May 2021.
[7] M. A. Hmida and A. Brahas, “An on-line condition monitoring system for incipient fault detection in double-cage induction motor,” IEEE Trans. Instrum. Meas., vol. 67, no. 8, pp. 1850–1858, Aug. 2018.
[8] Y. G. Lei and Z. J. He, “Advances in applications of hybrid intelligent fault diagnosis and prognosis technique,” J. Vib. Shock, vol. 30, no. 9, pp. 129–135, 2011.
[9] H. J. Wang, H. F. Wang, and Y. J. Ji, “Orbit identification method based on ISOMAP for rotor system fault diagnosis,” in Proc. IEEE 11th Int. Conf. Electron. Meas. Instrum., vol. 2, Aug. 2013, pp. 686–689, doi: 10.1109/ICIEME2013.6743131.
[10] S. Wang, X. Chen, G. Li, X. Li, and Z. He, “Matching demodulation transform with application to feature extraction of rotor rub-impact fault,” IEEE Trans. Instrum. Meas., vol. 63, no. 5, pp. 1372–1383, May 2014.
[11] W. L. Jiang, J. H. Li, Z. B. Li, and A. Q. Jiang, “Fault diagnosis method of rotor- shaft system based on the improved stacked denoising auto encoder depth model,” Mach. Tool Hydraulic, vol. 48, no. 21, pp. 182–188, 2020.
[12] M. Kim, J. H. Jung, J. U. Ko, H. B. Kong, J. Lee, and B. D. Yoon, “Direct connection-based convolutional neural network (DC-CNN) for fault diagnosis of rotor systems,” IEEE Access, vol. 8, pp. 172043–172056, 2020.
[13] Q. Ni, J. C. Ji, K. Feng, and B. Halkon, “A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis,” Mech. Syst. Signal Process., vol. 164, Feb. 2022, Art. no. 108216, doi: 10.1016/j.ymssp.2021.108216.
[14] R. Zhu and X. Xia, “Application of VMD in fault diagnosis for rotor-bearing system with rub-impact,” in Proc. IEEE 3rd Int. Conf. Inf. Commun. Signal Process. (ICICSP), Sep. 2020, pp. 502–505, doi: 10.1109/ICICSP50920.2020.9232055.
[15] J. L. Wu, X. N. Zhang, and B. Li, “A study on vibration component separation of a rotor system during startup and its application in fault diagnosis,” Meas. Sci. Technol., vol. 30, no. 9, 2019, Art. no. 095104, doi: 10.1088/1361-6501/ab199c.
[16] S. K. Liu and G. J. Tang, “Application of improved VMD method in fault diagnosis of rotor systems,” J. Chin. Soc. Power Eng., vol. 36, no. 6, pp. 448–453, 2016.
[17] X. Y. Zhang, X. L. Liu, and Z. Q. Luan, “Fault diagnosis of rotor based on VMD IWT approximate entropy and PSO_SVM,” Modular Mach. Tool Autom. Manuf. Techn., vol. 6, pp. 106–108 and 118, 2019, doi: 10.13462/j.cnki.mmtamt.2019.06.028.
[18] S. Liu, G. Tang, X. Wang, and Y. He, “Time-frequency analysis based on improved variational mode decomposition and teager energy operator for rotor system fault diagnosis,” Math. Problems Eng., vol. 2016, pp. 1–9, 2016.
[19] F. Wu, K. Cheng, J. Ding, and S. H. Liu, “Method for extracting fault feature information of multi-axis rotor system based on VMD-AR spectrum,” J. Jiangsu Univ. (Natural Sci. Ed.), vol. 41, no. 5, pp. 562–568, 2020.

[20] N. Lu, H. Xiao, and O. P. Malik, “Feature extraction using adaptive multiwavellets and synthetic detection index for rotor fault diagnosis of rotating machinery,” Mech. Syst. Signal Process., vol. 52–53, pp. 393–415, Feb. 2015, doi: 10.1109/JSSP.2014.07.024.

[21] H. Azami, M. Rostaghi, D. Abásolo, and J. Escudero, “Refined composite multiscale dispersion entropy and its application to biomedical signals,” IEEE Trans. Biomed. Eng., vol. 64, no. 12, pp. 2872–2879, Dec. 2017.

[22] H. Y. Zuo, X. B. Liu, and L. H. Hong, “Feature extraction and fault diagnosis based on FDM and RCMDE,” J. Vib. Meas. Diagn., vol. 41, no. 3, pp. 539–546, 2021.

[23] J. D. Wang, L. F. Ou, H. Y. Zhao, and M. P. Song, “Bearing fault diagnosis method for reciprocating compressor based on CEEMDAN and RCMDE,” Mach. Tool Hydraulic, vol. 49, no. 5, pp. 168–172+161, 2021.

[24] R. Gopalan, R. Li, and R. Chellappa, “Domain adaptation for object recognition: An unsupervised approach,” in Proc. Int. Conf. Comput. Vis., Nov. 2011, pp. 999–1006, doi: 10.1109/ICCV.2011.6126344.

[25] 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, doi: 10.1109/CVPR.2012.6247911.

[26] H. L. Cao, S. Gao, and P. Xue, “Aeroengine fault diagnosis based on multi-wavelets and synthetic detection index for rotor fault diagnosis of rotating machinery,” IEEE Trans. Biomed. Eng., vol. 68, pp. 1–16, 2020.

[27] J. Bjurgert, P. E. Valenzuela, and C. R. Rojas, “On adaptive boosting for machine fault diagnosis,” IEEE Access, vol. 7, pp. 26514–26526, 2019.

[28] C. Chen, F. Shen, and Y. R. Qiang, “Enhanced least squares square vector machine-based transfer learning strategy for bearing fault diagnosis,” Chin. J. Sci. Instrum., vol. 38, no. 1, pp. 33–40, Jan. 2017.

[29] H. L. Cao, S. Gao, and P. Xue, “Aeroengine fault diagnosis based on multi-classification AdaBoost,” J. Beijing Univ. Aeronaut. Astronaut., vol. 44, no. 9, pp. 1818–1825, 2018.

[30] G. Hao and B. Jiao, “Fault diagnosis of wind turbines’ bearing based on PSO-AdaBoostSVM,” in Proc. 3rd Int. Conf. Elect., Automat. Mech. Eng., vol. 127, 2018, pp. 252–255, doi: 10.2991/eame-18.2018.53.

[31] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” IEEE Trans. Signal Process., vol. 62, no. 3, pp. 531–544, Feb. 2014.

[32] S. Zhou, X. Xu, Z. Xu, W. Chang, and Y. Xiao, “Fractional-order modeling and fuzzy clustering of improved artificial bee colony algorithms,” IEEE Trans. Ind. Informat., vol. 15, no. 11, pp. 5988–5998, Nov. 2019.

[33] K. Li, R. Zhang, F. Li, L. Su, H. Wang, and P. Chen, “A new rotation machinery fault diagnosis method based on deep structure and sparse least squares support vector machine,” IEEE Access, vol. 7, pp. 26571–26580, 2019.

[34] J. Xue and B. Shen, “A novel swarm intelligence optimization approach: Sparrow search algorithm,” Syst. Sci. Control Eng., vol. 8, no. 1, pp. 22–34, Jan. 2020.

[35] T. Han, C. Liu, R. Wu, and D. Jiang, “Deep transfer learning with limited data for machinery fault diagnosis,” Appl. Soft Comput., vol. 103, May 2021, Art. no. 107150.

[36] Y. T. Liu and W. J. Sheng, “Fiber reinforced polymer damage identification based on AdaBoost-LSSVM,” China Meas. Test., vol. 46, no. 9, pp. 148–153, 2020.