Transfer Learning Based Data Feature Transfer for Fault Diagnosis

WEI XU [1], **YI WAN** [2,3], **(Member, IEEE), TIAN-YU ZUO** [1], **AND XIN-MEI SHA** [1]

[1] School of Mechanical and Electrical Engineering, SanJiang University, Nanjing 210012, China
[2] School of Environmental Science, Nanjing Xiaozhuang University, Nanjing 211171, China
[3] Jiangsu Shentong Valve Company, Ltd., Nantong 226232, China

Corresponding author: Yi Wan (wan_sju@163.com)

ABSTRACT The development of sensor technology provides massive data for data-driven fault diagnosis. In recent years, more and more scholars are studying artificial intelligence technology to solve the bottleneck in fault diagnosis. Compared with other classification and prediction problems, fault diagnosis often faces the problem of data scarcity. To overcome the lack of fault data, the transfer learning based on different working condition is gradually introduced into fault diagnosis by scholars. This paper discusses the current mainstream AI-based fault diagnosis methods, and analyzes the advantage of transfer learning for fault diagnosis problem. Then, a transfer component analysis (TCA) based method is proposed to transfer data features between different working conditions. Through the TCA-based method, the fault diagnosis model under the working condition can be established with the help of historical working condition. It effectively alleviates the problem of data scarcity under the condition to be predicted. Different from other fault diagnosis studies, this paper considers the online maintenance process based on TCA. A fault diagnosis framework including online maintenance process is proposed. Finally, a case study of bearing diagnosis from Case Western Reserve University proves the feasibility and effectiveness of the proposed TCA-based method and our fault diagnosis framework.

INDEX TERMS Fault diagnosis, feature extraction, feature transfer, sensors.

I. INTRODUCTION

With the popularization of sensor technology, more and more equipment in industrial manufacturing has achieved effective digital monitoring [1], [2]. Through digital monitoring, it is possible to quantify the operating status of the equipment, and comprehensively evaluate the health status of the equipment, thereby improving the reliability of the equipment and ensuring the quality of product manufacturing.

A complete fault diagnosis process generally includes two steps: fault signal processing and fault pattern recognition. Based on digital monitoring, academia has carried out extensive research work on fault signal processing [3]–[5] and fault pattern recognition [6]–[8], respectively. Fault signal processing mainly focuses on the separation and extraction of fault features to guide the subsequent fault recognition work. Fault pattern recognition often focuses on how to identify the

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

type of fault by matching fault features obtained by signal processing with existing faults in the system.

On the one hand, high-precision pattern recognition and classification depends on the effective description and characterization of signals. In other words, an effective signal description will provide highly distinguished signal features for pattern recognition, thereby improving the accuracy of pattern recognition. On the other hand, the validity of signal characterization needs to be verified by pattern recognition. Recently, many scholars have proposed a large number of indicators and methods for signal characterization from the perspective of time domain and frequency domain. However, for different types of signals, it is still necessary to combine subsequent pattern recognition to verify the reasonableness of signal feature extraction. It is worth noting that some scholars have begun to advocate the application of deep neural networks and self-encoding methods to combine fault signal processing and fault pattern recognition together for fault diagnosis [7], [9]–[13]. The introduction of the deep neural

network method and the comparison between this method and the traditional method will be analyzed in the Section II.

Since fault diagnosis is mainly conducted by data-driven methods, data play a vital role in the effectiveness of fault pattern recognition. But it is well known that the number of fault samples is small compared to normal samples. The problem of data class imbalance will cause many data-sensitive classification models to be overfitting [14]. Even in extreme cases, the classification model will classify each sample as the normal one if they ignore those few fault samples [15], [16]. In this case, although the classification accuracy of the model can still approach almost 100%, the meaning of classification has been lost.

Theoretically, data class imbalance can be solved by cost-sensitive weighting [15]. But for equipment fault diagnosis in the actual industry, cost-sensitive weighting is difficult to set. Therefore, more attention should be paid to the data source to solve this problem, that is, how to obtain as many and comprehensive types of fault data as possible. The easiest way to get the fault characteristics data is to simulate the fault scenario. Unfortunately, for most equipment or components, it is not feasible to deliberately destroy them to obtain fault features data. On the one hand, in general, one equipment will correspond to multiple faults and simulation experiments require a lot of resource support. On the other hand, since the equipment operates in specific scenarios, the fault characteristics under different working conditions are also different. Therefore, it is difficult to characterize fault features by exhausting all working conditions through simulation experiments.

In industrial practice, in addition to large-scale experiments, it is more likely to obtain fault characteristics from the following two situations.

(1) We only have a small amount of historical fault data for equipment that needs to be diagnosed, and we also have some fault data for several equipment similar to this equipment. Their functions are similar with minor differences in parameters. Therefore, the faults between them have similar characteristics and can be used to predict the fault types for each other.

(2) For one equipment, the equipment runs a lot of time under a certain operating condition, so a large amount of fault data are accumulated. When the equipment is put into a similar working condition, the fault characteristics also will be similar. Therefore, the fault types under the new working condition can be predicted by these fault types under the old working condition.

In view of the above circumstances, this paper adopts feature transfer to solve the problems of low classification accuracy and invalid classification caused by insufficient fault data. Among algorithms of feature transfer, TCA is a typical method for domain adaptation problems. It maps the both data of the source domain and target domain to a high-dimensional Reproducing Kernel Hilbert Space (RKHS). In RKHS, the data distance between the source and

target domain is minimized, while retaining their respective internal attributes to the greatest extent.

The contribution of this paper is described as follows. The extraction method of data features based on TCA is proposed to discover general features between source domain and target domain. Then, different from other researches, this work further explores how to use few data labels from target domain to improve the accuracy of target domain fault diagnosis. In the actual maintenance work, the occurrence of fault in the target domain is gradual (as the maintenance continues, the maintenance personnel continue to label the fault data in the target domain). Therefore, how to add the newly acquired data in the target domain to the fault diagnosis will be a meaningful work.

The remainder of this paper is organized as follows. Section II reviews and compares the current mainstream methods of fault signal processing, fault pattern recognition and fault diagnosis. The feature transfer method based on transfer component analysis is introduced in Section III to support subsequent fault diagnosis. A case study is given to verify the effectiveness of the proposed method in Section IV. Finally, the conclusion is arranged in Section V.

II. RELATED WORK
A. FAULT DIAGNOSIS BASED ON SIGNAL PROCESSING
Signal processing-based fault diagnosis usually consists of signal feature extraction and fault pattern recognition. The following describes the current researches on signal feature extraction and fault pattern recognition.

1) SIGNAL FEATURE EXTRACTION
The feature extraction of fault signals is to extract statistical feature values for samples with a specific length. Signal feature extraction can be mainly extracted from time domain, frequency domain and time- frequency domain [17]–[20].

The time domain indicator includes Mean, Standard Deviation, Variance, Peak-Peak Value, Root Mean Square, Waveform Factor, Crest Factor, Impulse Factor, Kurtosis [21], [22], Skewness [22] and etc.

The frequency domain indicator includes Frequency Centre (FC), Root Mean Square Frequency (RMSF), Root Variance Frequency (RVF) [17], and etc.

The time- frequency domain indicator includes Short-Time Fourier Transform (STFT) [23]–[25], Wavelet Transform [26], [27], Empirical Mode Decomposition [28], [29], Wigner-Ville Distribution (WVD) [30], [31] and etc.

2) FAULT PATTERN RECOGNITION
As we mentioned in Section I, the input of fault pattern recognition is the output of signal processing. Through the signal processing in the previous step, the fault signal can be converted into multiple fault samples. For samples with known fault types, learning algorithms can be used to train the classification model, then to identify and predict fault pattern.

In recent years, various machine learning algorithms have become the mainstream methods for fault pattern recognition. For example, [32] conducted vibration fault recognition for hydroelectric generating via support vector machine. Reference [33] used decision tree for spur gear fault diagnostics. Bayesian Network was used to Intelligent Building Fault Diagnosis in [34]. Moreover, machine learning-based pattern recognition is often combined with computational intelligence methods such as evolutionary computation and fuzzy systems [35], [36]. For example, some researches used algorithms such as PSO and ACO to optimize model parameters for machine learning, thereby improving recognition accuracy [32], [37]. There are also some works that combine methods such as fuzzy decision making and information fusion with machine learning to improve the robustness of the pattern classifier [33], [38].

B. FAULT DIAGNOSIS BASED ON DEEP NEURAL NETWORK
In recent years, more and more researches have discussed how to use deep neural networks for fault diagnosis [7], [39]–[46]. Neural network-based fault diagnosis can integrate signal feature extraction and fault pattern recognition. This type of research work extracts signal features and fault mode recognition, directly imports the signals as samples into deep neural networks, uses the network to extract signal features, and directly derives the pattern classification results at the last layer of the network. For example, [47], [48] discussed how to use deep neural network with autoencoders for fault diagnosis. The first half of the neural network is responsible for signal extraction and the second half is responsible for pattern recognition. The advantage of such fault diagnosis method is that it can save the cost of manual signals extraction and use the network training to perform feature extraction autonomously.

However, such a fault diagnosis method has at least two obvious shortcomings as follows.
- First, Difficulties in network training.

For neural networks, if the data samples and feature sizes are relatively large, then it will be difficult to use neural networks for training. On the one hand, network training takes a long time, so that it is difficult to quickly put the model into use. On the other hand, due to the huge scale of the network, training convergence is difficult to be guaranteed.
- Second, poor interpretability.

Due to the black box training mode of neural network, the interpretability of the parameters after training is very poor, and the meaning of the fault features cannot be obtained intuitively and clearly. When the scale of the training network is large, even if high recognition accuracy is obtained, it is difficult to judge whether the model has really learned the features or the high accuracy is caused by overfitting [14].

C. DISCUSSION ON FAULT DIAGNOSIS METHOD
Intelligent fault diagnosis is different from artificial intelligence applications. The interpretability and effectiveness of its prediction results determine whether industrial activities

can proceed smoothly and even affect important industrial safety issues. Therefore, although neural networks can make some progress in this filed, this paper still advocates using signal processing to extract features, which can retain the interpretability of the fault characteristics to the greatest extent, and it is also convenient for the fault maintenance personnel to evaluate the results of fault pattern recognition.

D. TRANSFER LEARNING FOR FAULT PATTERN RECOGNITION
The research on fault transfer benefits from the development of transfer learning [49]. The mainstream researches on fault transfer includes the following three categories.
- (1) Transfer from Data

Due to the scarcity of fault data in a specific working condition, if fault data similar to the working condition are known, these data can be directly used as auxiliary data. By introducing auxiliary data into the current working condition, the fault classification model can be established, which can be used to predict the potential fault in the unknown signal.
- (2) Transfer from Model and Parameter [50]

Unlike data transfer, model Transfer is performed from the perspective of algorithm and model. In other words, the structure and parameters of the classification model established under known work condition can be transferred to the new work conditions. By fine-tuning the structure and parameters [51], the model established under known work condition can be easy to classify and predict faults under new work condition.
- (3) Transfer from Feature

Feature transfer focuses on the relationship between auxiliary data and the data generated by existing working condition. In general, the auxiliary data generated from past known working condition is large, and after a period of operation, the sensor system will also obtain some fault data generated under existing working condition. Therefore, data feature matching between the both data can be used to classify and predict based on the similarity of features.

All three above transfer strategies, to a certain extent, help alleviate the problem of sparse fault data under specific working condition. From the perspective of industrial applications, not all three are effective.

Data transfer is mainly used in situations where data are generated by approximate data distribution. Once the data distribution differs greatly under approximate working condition, it will cause a huge prediction error in predictive diagnosis.

The transfer of model and parameter is generally carried out on neural networks. Compared to data transfer, this method avoids the distribution error caused by the direct use of auxiliary data. And, from the perspective of model training, it does not need to train all parameters of the entire model, but only needs do some fine-tuning based on the existing model. However, the shortcoming of this method is also obvious, that is, the method has a strong dependence on the model.

Once the model selection or model parameter fine-tuning is not done well, the effect of the entire model will be very poor.

Feature transfer is a method that can combine the feature of auxiliary data with the signal of the current working condition. This method considers the difference of data under different working condition and advocates a simpler fault prediction model. The process is as follows: In the first step, the both signal from historical known working condition and the current working condition have undergone spatial transformation simultaneously, which significantly reduces the difference between auxiliary data and current target data. Then, in the second step, the transformed auxiliary data and the predicted working condition data are combined into a new sample. Due to the elimination of differences, a simpler model can be used for fault classification and prediction, which reduces the dependence of the classification on the model.

Based on the above advantages of feature transfer, this paper advocates using feature transfer to transfer fault data under different working conditions. The specific modeling of fault diagnosis based on feature transfer will be given in the subsequent Section III.

III. FAULT DIAGNOSIS BASED ON FEATURE TRANSFER
A. MODEL ASSUMPTIONS
Assume the auxiliary fault data under known conditions is D_S, which is defined in Equation (1).

$$D_S = [X_S, y_S] = \begin{bmatrix} x_S^{1,1} & \cdots & x_S^{1,d} & y_S^1 \\ x_S^{2,1} & \cdots & x_S^{2,d} & y_S^2 \\ \vdots & \ddots & \cdots & \cdots \\ x_S^{n_s,1} & \cdots & x_S^{n_s,d} & y_S^{n_s} \end{bmatrix} \quad (1)$$

where X_S represents the extracted fault signal features, d represents the dimension of the fault signal features, n_s represents the number of samples under known conditions, and y_S represents the fault type corresponding to the fault signal.

The fault data under the working condition to be diagnosed is D_T, which is defined in Equation (2).

$$D_T = [X_T] = \begin{bmatrix} x_T^{1,1} & x_T^{1,2} & \cdots & x_T^{1,d} \\ x_T^{2,1} & x_T^{2,2} & & x_T^{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_T^{n_T,1} & x_T^{n_T,2} & \cdots & x_T^{n_T,d} \end{bmatrix} \quad (2)$$

where X_T represents the extracted fault signal features, d represents the dimension of the fault signal features (consistent with the dimension of X_S), and n_T represents the number of samples under the working condition to be diagnosed.

The notation y_T is used to indicate the fault type of the fault data under the working condition to be diagnosed.

B. FEATURE TRANSFER BASED ON TRANSFER COMPONENT ANALYSIS
Feature transfer is performed by transfer component analysis (TCA) [52]. Maximum mean discrepancy (MMD) is used to characterize the difference between the distribution of auxiliary data and the data to be diagnosed in TCA method. MMD is given by Equation (3).

$$dist(X_S, X_T) = \left\| \frac{1}{n_S} \sum_{i=1}^{n_S} \emptyset(X_S) - \frac{1}{n_T} \sum_{i=1}^{n_T} \emptyset(X_T) \right\| \quad (3)$$

That is, through mapping transformation $\emptyset$, X_S and X_T can be mapped to a high-dimensional space, and in this space X_S and X_T are as close as possible.

Considering the difficulty of solving the mapping transformation $\emptyset$, a kernel matrix K is introduced which is given by Equation (4).

$$K = \begin{bmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{bmatrix} \quad (4)$$

And parameter matrix L is defined by Equation (5).

$$L_{ij} = \begin{cases} \dfrac{1}{n_S^2}, & x_i, x_j \in X_S, \\ \dfrac{1}{n_T^2}, & x_i, x_j \in X_T, \\ -\dfrac{1}{n_S n_T}, & otherwise. \end{cases} \quad (5)$$

Now, the original problem can be transformed into the solution of K in the following Equation (6).

$$trace(KL) - \lambda trace(K) \quad (6)$$

The problem can still be further structured by dimension reduction.

$$\tilde{K} = \left(KK^{-\frac{1}{2}} \tilde{W} \right) \left(\tilde{W}^T K^{-\frac{1}{2}} K \right) = KWW^T K \quad (7)$$

The matrix W here is a matrix with a lower dimension than K. As long as W is solved, the original problem can be solved.

Sort out the final optimization goals of TCA with W, which can be formulated by Equation (8).

$$\min_W tr\left(W^T KLKW \right) + \mu tr\left(W^T W \right)$$
$$s.t.\ W^T KHKW = I_m \quad (8)$$

where $tr\left(W^T W \right)$ is norm penalty for parameter [53] and H is a center matrix defined by Equation (9).

$$H = I_{n_S + n_T} - \frac{1}{n_S + n_T} \mathbf{1}\mathbf{1}^T \quad (9)$$

The optimization goal can be obtained by Lagrangian duality [54]. Let D be the feature vector corresponding to the maximum m eigenvalues of $(KLK + \mu I)^{-1} KHK$, then there is a relationship among D, $\emptyset(X_S)$, and $\emptyset(X_T)$, shown in Equation (10).

$$D = \begin{bmatrix} \emptyset(X_S) \\ \emptyset(X_T) \end{bmatrix} \quad (10)$$

In this way, the features X_S and X_T after transformation can are obtained. Where m determines the transformed feature dimension. Similar to the Principal Component Analysis [55], the larger m, the closer the transformed features are.

C. FAULT DIAGNOSIS BASED ON FEATURE TRANSFER

Once the distance between auxiliary data and the current target data is significantly reduced in the high-dimensional space Ø, the auxiliary data can be used to build a fault classification model. The classification model here can be a neural network, a support vector machine, a decision tree or other learning algorithms. In this paper, we use k-Nearest Neighbor (KNN) to do the classification work.
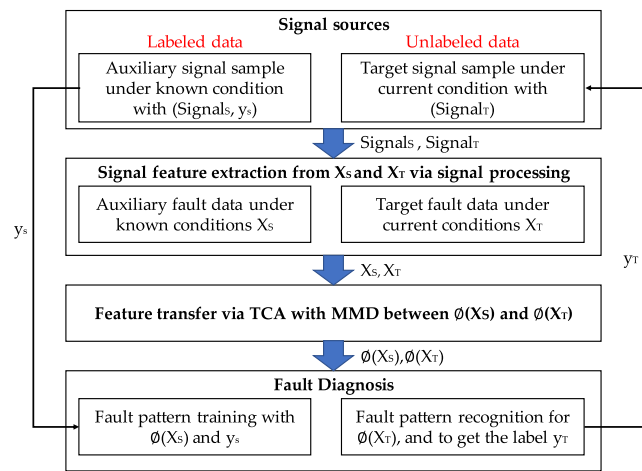

FIGURE 1. The framework of fault diagnosis based on feature transfer.

Figure 1 shows the framework of fault diagnosis based on feature transfer. What needs to be explained in Figure 1 is that, for fault diagnosis itself, fault data and fault discrimination results are constantly generated. Therefore, on the one hand, it is possible to continuously generate new fault data into the TCA and iteratively update the results of feature transfer. On the other hand, maintenance personnel will continue to label fault data under current working condition, which could be added into the training of fault classification model for a better fault diagnosis result.

From the perspective of algorithm, our framework of fault diagnosis adopts TCA to predict the fault type in target domain. The advantage of TCA is that the implementation is simple. The method itself does not have too many limitations, and it is as easy to use as Principal Component Analysis (PCA). More importantly, compared with the traditional transfer learning framework, our framework also takes the online maintenance of the target domain into account. In other words, the framework allows the labeled fault data obtained in the target domain to be added to the training of the fault diagnosis model in real time. It will help to further improve the accuracy of fault diagnosis in the target domain.

IV. CASE STUDY
A. DATA DESCRIPTION

The data used in this paper come from the bearing data center of Case Western Reserve University (CWRU). CWRU provides test data of different size bearings and corresponding bearing faults. In CWRU experiments, the bearing is installed on the 2 horsepower (hp) motor. The real-time acceleration data are measured at different sampling frequencies (12K and 48K) by sensors installed near and far from the motor bearing.
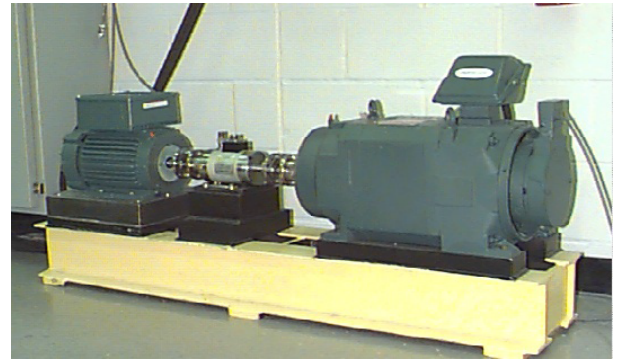

FIGURE 2. Experimental equipment for bearing failure.

Figure 2 shows the basic structure of the test environment for bearing fault, which is composed of 2 hp motor (left), torque sensor/encoder (Center), dynamometer (right) and control electronics (not shown). The fault diameter of bearing is 0.007," 0.014," 0.021," 0.028" and 0.040," where 0.007" means 0.007 inch. This paper mainly studies on 0.007," 0.014," 0.021," and 0.028." In order to simulate the difference of fault characteristics under different working conditions, motor loads are composed of four level, including 0, 1, 2, 3 (corresponding motor speeds from 1797 to 1720 rpm).

TABLE 1. The working conditions used in the case study.

Motor load (hp)	Hp_1	Hp_2	Hp_3	Hp_4
Working Condition (rpm)	1797	1772	1750	1730

TABLE 2. The fault diameter and the corresponding fault types ID used in the case study.

Rolling bearing condition		Class Label
Fault Diameter	Fault Type	
0.007"	Inner Race	1
0.007"	Ball	2
0.007"	Outer Race	3
0.014"	Inner Race	4
0.014"	Ball	5
0.014"	Outer Race	6
0.021"	Inner Race	7
0.021"	Ball	8
0.021"	Outer Race	9

For clear expression of the experimental data, Table 1 shows the working conditions used in our case study, and Table 2 shows the fault diameter and the corresponding fault

(a) Normal Data of 3200 Points at Rpm 1772 Under 48K

(b) 0.007" Inner Race Fault Data of 800 Points at Rpm 1772 Under 12K

(C) 0.014" Inner Race Fault Data of 800 Points at Rpm 1772 Under 12K

(d) 0.007" Ball Fault Data of 800 Points at Rpm 1772 Under 12K

FIGURE 3. Four groups of time series signals at speed 1772.

types ID (for example, if there is a fault of 0.007'' in the inner race of the bearing, the fault types ID will be called as '1'). In addition, there are also normal data in the experiment. We will use '0' as the type ID.

B. FEATURE EXTRACTION FROM FAULT DATA

Figure 3 shows four groups of time series signals at speed 1772 rpm. Since the normal data in CRWU experiment are collected at 48K, compared to 12K, its data points are 4 times

than that of 12K. For intuitive presentation, the time series signal of the normal data gives 3200 points of the time series signal of the normal data are shown in sub-figure(a), and 800 points of fault data are shown in sub-figure (b) (c) (d). It can be clearly seen from the signal comparison that the data from normal operation have a lower value change range and the signal is smoother than others. From the comparison of (b) and (c), it can be seen that the same fault type under different fault diameters has similar change patterns but different amplitudes. Therefore, this situation has the possibility of feature transfer. From the comparison of (b) and (d), it shows that there are obvious differences in the signal changes corresponding to different fault types, so that it is possible to be classified.

TABLE 3. Time domain indicators used in the case study.

Indicator	Formulation
Peek to Peek Value (P2P)	$max(x(i)) - min(x(i))$
Root Mean Square (RMS)	$\sqrt{\dfrac{1}{N}\sum_{i=1}^{N} x^2(i)}$
Waveform Factor (WF)	$\dfrac{RMS}{mean(abs(x(i)))}$
Crest Factor (CF)	$\dfrac{max(|x(i)|)}{RMS}$
Impulse Factor (IF)	$\dfrac{max(|x(i)|)}{mean(abs(x(i)))}$
Kurtosis (K)	$\dfrac{\frac{1}{N}\sum_{i=1}^{N}(x(i)-\mu)^4}{S^4}$
Skewness (S)	$\dfrac{1}{N}\sum_{i=1}^{N}(x(i)-\mu)^3$

Through the comparison of Figure 3, we find that there are obvious differences in the time series signals of different faults, so the time domain indicators of the signals are used to extract the features of the signals. Therefore, here we mainly consider the following seven time-domain indicators in Table 3 for feature extraction. And, set the time window scale of feature extraction with every 800 points in the 12K as a sample, and every 3200 points in the 48K as a sample. Use the above time-domain indicators to process all samples, and regularize all features to ensure that the scales between different indicators are consistent.

C. FAULT RECOGNITION
1) FAULT RECOGNITION UNDER 12K
After feature extraction, for 12K, 150 samples are obtained for each fault type and the normal operation. At 12K, the accuracy of fault prediction after feature transfer for the case of 0.007'' and 0.021'' is tested respectively. Here, each data set includes three fault classes and one normal class (four-classes fault recognition).

Table 4 gives the transfer results among four types of speeds (See Table 1) one by one under the case of 0.007.'' For example, the result from Row 2, Column 3 means the accuracy of the data feature from Hp_2 to Hp_3.

TABLE 4. The transfer results among different speeds in the case of 0.007'' under 12K.

	Hp_1	Hp_2	Hp_3	Hp_4
Hp_1	-	0.965	0.987	0.970
Hp_2	0.995	-	0.980	0.972
Hp_3	0.985	0.977	-	0.967
Hp_4	0.993	0.948	0.978	-

TABLE 5. The transfer results among different speeds in the case of 0.021'' under 12K.

	Hp_1	Hp_2	Hp_3	Hp_4
Hp_1	-	0.998	0.993	1.000
Hp_2	1.000	-	0.993	1.000
Hp_3	0.993	0.985	-	1.000
Hp_4	0.977	0.978	0.998	-

Same as Table 4, Table 5 gives the results under the case of 0.021.''

It can be found from Tables 4 and 5 that although the feature transfer uses only seven time-domain indicators, its prediction accuracy is quite high, proving the effectiveness of the data feature transfer.

2) FAULT RECOGNITION UNDER 48K
Since the amount of data at 48K is more abundant, we test more types of fault predictions in the 48K. Here we consider six types of faults 1, 2, 4, 5, 7, 8 and normal operation 0. That is, consider a seven-classes fault recognition, which is more difficult than four-classes fault recognition problem under 12K. The fault recognition accuracy is tested at three speeds: Hp_2, Hp_3, Hp_4 and the results are shown in the following Table 6.

TABLE 6. The transfer results among different speeds under 48K.

	Hp_2	Hp_3	Hp_4
Hp_2	-	0.995	0.882
Hp_3	0.915	-	0.995
Hp_4	0.925	0.998	-

It can be found that its classification accuracy is still very good. Therefore, the validity of feature transfer on fault recognition is further proved.

According to Figure 1, the predicted results under the current working condition can be compared with the actual fault after inspection by maintenance personnel, then the correct label can be obtained. By using these correct labeled data under the current working condition and auxiliary data under the historical similar working condition for training, the accuracy of the model can also be improved. Now, assume that 10% of the data have been labeled for the current working condition. That is, for original 150 samples of each class, 15 samples have been labeled by maintenance personnel. Table 7 gives the transfer results among different speeds under 48K after 10% samples being labeled.

TABLE 7. The transfer results among different speeds under 48K after 10% samples being labeled.

	Hp_2	Hp_3	Hp_4
Hp_2	-	1.000	0.956
Hp_3	0.985	-	0.998
Hp_4	0.975	0.998	-

It can be found that this approach is feasible, and almost all the results have been significantly improved than the original. The rationality and effectiveness of our method are verified again by compared with Table 6.

D. COMPARISON WITH OTHER ALGORITHMS

To verify the effectiveness of our TCA based method, deep belief network (DBN), support vector machine (SVM), artificial neural network (ANN), Bayesian method, Bagging method and Boosting method are used to compare with the proposed method. The accuracy results of the above methods for fault prediction have been given in [56]. Our proposed method uses KNN for classification after TCA. In order to illustrate the effectiveness of TCA for classification, KNN is used as beachmark in following experiment.

The comparison experiments are conducted on two data set, termed Group A and Group B. Group A represents transferring fault predict model from Hp_1 to Hp_2. Group B represents transferring the fault predict model from Hp_3 to Hp_4. Both data set all come from sampling under 12K. Different from Section IV.C, comparison experiments in this section consider all fault types (Inner Race, Ball, Outer Race) and all fault diameter (0.007," 0014" and 0.021"). Therefore, there are nine fault classes and one normal class (i.e., ten-classes fault recognition).

To eliminate the interference caused by different data, we use the same method in [56] to generate the experiment data. For each fault class and normal class, 300 samples are generated for training, and 200 samples are used for testing.

For a clearer expression, Table 8 gives the experiment data settings.

TABLE 8. Description of group A and group B.

Rolling bearing condition	Group A Training (Hp_1)/ Testing (Hp_2)	Group B Training (Hp_3)/ Testing (Hp_4)	Class Label
Normal	300/200	300/200	0
0.007"- Inner Race	300/200	300/200	1
0.007"- Ball	300/200	300/200	2
0.007"- Outer Race	300/200	300/200	3
0.014"- Inner Race	300/200	300/200	4
0.014"- Ball	300/200	300/200	5
0.014"- Outer Race	300/200	300/200	6
0.021"- Inner Race	300/200	300/200	7
0.021"- Ball	300/200	300/200	8
0.021"- Outer Race	300/200	300/200	9

Table 9 gives the comparison results of different methods. It can be clearly seen that the proposed TCA based method

TABLE 9. The transfer results of different algorithms on group A and group B.

Methods	Group A	Group B
TCA based method	**0.892**	**0.914**
KNN	0.844	0.670
DBN	0.867	0.882
SVM	0.562	0.596
ANN	0.673	0.681
Bayesian	0.561	0.605
Bagging	0.442	0.491
Boosting	0.556	0.582

outperforms other methods on both test groups. It can be seen that the benchmark results (KNN results) are quite different on Group A and B. The TCA based method can obtain good results on both test groups, especially for Group B. It shows that TCA has indeed learned the fault features between data from source domain and target domain.

V. CONCLUSION

This paper studies data-driven fault diagnosis, and analyzes the difference between classical signal processing-based fault diagnosis and deep learning fusion diagnosis. It is clear that the method using signal processing flow in fault diagnosis has better results and robustness, and can be promoted into various industrial fault diagnosis problems. Then, through the combination of fault diagnosis process and data feature transfer method, a fault diagnosis framework is proposed. Then, the framework is used in the fault diagnosis of CWRU bearing data, which proves the validity and rationality of the proposed framework and corresponding methods.

At present, the mainstream transfer learning methods require a high degree of similarity between different data domains. It restricts the use of transfer learning in fault diagnosis. Therefore, in future research, our research will focus on the fault transfer in case of great difference of working conditions. And, traditional transfer learning in fault diagnosis is always one-to-one (i.e., one source domain to one target domain). In many practical cases, there may be more than one historical working conditions. How to use the data from different historical working conditions to assist the fault diagnosis under the target working condition will be an urgent research work in the future.

REFERENCES

[1] M. G. Pecht and M. Kang, "Sensor systems for PHM," in *Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things*. Piscataway, NJ, USA: IEEE Press, 2019, pp. 39–60, doi: 10.1002/9781119515326.ch2.

[2] V. Holopov, A. Kushnir, E. Kurnasov, A. Ganichev, and A. Romanov, "Development of digital production engineering monitoring system based on equipment state index," in *Proc. IEEE Conf. Russian Young Researchers Electr. Electron. Eng. (EIConRus)*, Feb. 2017, pp. 863–868.

[3] A. Rai and S. H. Upadhyay, "A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings," *Tribology Int.*, vol. 96, pp. 289–306, Apr. 2016.

[4] S. Lu, Q. He, H. Zhang, and F. Kong, "Rotating machine fault diagnosis through enhanced stochastic resonance by full-wave signal construction," *Mech. Syst. Signal Process.*, vol. 85, pp. 82–97, Feb. 2017.

[5] Z. Qiao, Y. Lei, J. Lin, and F. Jia, "An adaptive unsaturated bistable stochastic resonance method and its application in mechanical fault diagnosis," *Mech. Syst. Signal Process.*, vol. 84, pp. 731–746, Feb. 2017.

[6] A. Glowacz, "Fault diagnosis of single-phase induction motor based on acoustic signals," *Mech. Syst. Signal Process.*, vol. 117, pp. 65–80, Feb. 2019.

[7] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2016.

[8] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, Jan. 2017.

[9] Z. Chen, S. Deng, X. Chen, C. Li, R.-V. Sanchez, and H. Qin, "Deep neural networks-based rolling bearing fault diagnosis," *Microelectron. Rel.*, vol. 75, pp. 327–333, Aug. 2017.

[10] J. J. Q. Yu, Y. Hou, A. Y. S. Lam, and V. O. K. Li, "Intelligent fault detection scheme for microgrids with wavelet-based deep neural networks," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1694–1703, Mar. 2019.

[11] R. Zhang, Z. Peng, L. Wu, B. Yao, and Y. Guan, "Fault diagnosis from raw sensor data using deep neural networks considering temporal coherence," *Sensors*, vol. 17, no. 3, p. 549, 2017.

[12] Z. Wang, J. Wang, and Y. Wang, "An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition," *Neurocomputing*, vol. 310, pp. 213–222, Oct. 2018.

[13] J. Wang, S. Li, Z. An, X. Jiang, W. Qian, and S. Ji, "Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines," *Neurocomputing*, vol. 329, pp. 53–65, Feb. 2019.

[14] D. M. Hawkins, "The problem of overfitting," *J. Chem. Inf. Comput. Sci.*, vol. 44, no. 1, pp. 1–12, Jan. 2004.

[15] B. Zadrozny, J. Langford, and N. Abe, "Cost-sensitive learning by cost-proportionate example weighting," in *Proc. 3rd IEEE Int. Conf. Data Mining*, Nov. 2003, pp. 435–442.

[16] C. X. Ling and V. S. Sheng, "Cost-sensitive learning and the class imbalance problem," *Encyclopedia of Mach. Learn.*, vol. 2011, pp. 231–235, Jan. 2008.

[17] W. Caesarendra and T. Tjahjowidodo, "A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing," *Machines*, vol. 5, no. 4, p. 21, 2017.

[18] M. Žvokelj, S. Zupan, and I. Prebil, "EEMD-based multiscale ICA method for slewing bearing fault detection and diagnosis," *J. Sound Vibrat.*, vol. 370, pp. 394–423, May 2016.

[19] C. Lu, J. Chen, R. Hong, Y. Feng, and Y. Li, "Degradation trend estimation of slewing bearing based on LSSVM model," *Mech. Syst. Signal Process.*, vols. 76–77, pp. 353–366, Aug. 2016.

[20] H. Wang, Y. Tang, and R. Hong, "Multiple physical signals based residual life prediction model of slewing bearing," *J. Vibroeng.*, vol. 18, no. 7, pp. 4340–4353, Nov. 2016.

[21] Y. Wang, J. Xiang, R. Markert, and M. Liang, "Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications," *Mech. Syst. Signal Process.*, vols. 66–67, pp. 679–698, Jan. 2016.

[22] K. V. Mardia, "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, vol. 57, no. 3, pp. 519–530, 1970.

[23] A. Widodo and B. Yang, "Wavelet support vector machine for induction machine fault diagnosis based on transient current signal," *Expert Syst. Appl.*, vol. 35, nos. 1–2, pp. 307–316, Jul. 2008.

[24] Z. Feng, M. Liang, and F. Chu, "Recent advances in time-frequency analysis methods for machinery fault diagnosis: A review with application examples," *Mech. Syst. Signal Process.*, vol. 38, pp. 165–205, Jul. 2013.

[25] K. Satpathi, Y. M. Yeap, A. Ukil, and N. Geddada, "Short-time Fourier transform based transient analysis of VSC interfaced Point-to-Point DC system," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4080–4091, May 2018.

[26] L.-H. Wang, X.-P. Zhao, J.-X. Wu, Y.-Y. Xie, and Y.-H. Zhang, "Motor fault diagnosis based on short-time Fourier transform and convolutional neural network," *Chin. J. Mech. Eng.*, vol. 30, no. 6, pp. 1357–1368, Nov. 2017.

[27] J. Chen, J. Pan, Z. Li, Y. Zi, and X. Chen, "Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals," *Renew. Energy*, vol. 89, pp. 80–92, Apr. 2016.

[28] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 112–114, Feb. 2004.

[29] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. J. Amaratunga, "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting," *Appl. Soft Comput.*, vol. 54, pp. 246–255, May 2017.

[30] Z. Liu, K. Wu, Z. Ma, and Q. Ding, "Vibration analysis of a rotating Flywheel/Flexible coupling system with angular misalignment and rubbing using smoothed pseudo Wigner–Ville distributions," *J. Vibrat. Eng. Technol.*, pp. 1–12, Nov. 2019, doi: 10.1007/s42417-019-00189-y.

[31] J.-H. Cai and Q.-Y. Chen, "Bearing fault diagnosis method based on local mean decomposition and wigner higher moment spectrum," *Experim. Techn.*, vol. 40, no. 5, pp. 1437–1446, Oct. 2016.

[32] J. Luo, X. Wang, and Y. Xu, "Vibration fault diagnosis for hydro-electric generating unit based on generalized S-transform and QPSO-SVM," in *Proc. IEEE Sustain. Power Energy Conf. (iSPEC)*, Nov. 2019, pp. 2133–2137.

[33] A. Krishnakumari, A. Elayaperumal, M. Saravanan, and C. Arvindan, "Fault diagnostics of spur gear using decision tree and fuzzy classifier," *Int. J. Adv. Manuf. Technol.*, vol. 89, nos. 9–12, pp. 3487–3494, Apr. 2017.

[34] J. Fu, L. Huang, and L. Chen, "Intelligent building fault diagnosis based on wavelet transform and Bayesian network," in *Proc. Int. Conf. Smart City Intell. Building*. Singapore: Springer, 2018, pp. 419–432.

[35] J. K. De, "Evolutionary computation: A unified approach," in *Proc. Genetic Evol. Comput. Conf. Companion*, Jul. 2016, pp. 185–199.

[36] X. Sun and X. Jia, "A fault diagnosis method of industrial robot rolling bearing based on data driven and random intuitive fuzzy decision," *IEEE Access*, vol. 7, pp. 148764–148770, 2019.

[37] J. Wang, G. Feng, C. Yi, F. Xu, Y. Wang, and Z. Liu, "Electronic circuit fault diagnosis based on an ACO-RVM classifier with a new nonlinearly mixed kernel function," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Paris)*, May 2019, pp. 332–337.

[38] Z.-H. Sun, R. Liu, and X. Ming, "A fault diagnosis and maintenance decision system for production line based on human-machine Multi- information fusion," in *Proc. Artif. Intell. Cloud Comput. Conf. ZZZ AICCC*, 2018, pp. 151–156.

[39] W. Sun, S. Shao, and R. Yan, "Induction motor fault diagnosis based on deep neural network of sparse auto-encoder," *J. Mech. Eng.*, vol. 52, no. 9, p. 65, 2016.

[40] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng, and W. Jiang, "Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks," *Neurocomputing*, vol. 315, pp. 412–424, Nov. 2018.

[41] Y. Li, G. Cheng, C. Liu, and X. Chen, "Study on planetary gear fault diagnosis based on variational mode decomposition and deep neural networks," *Measurement*, vol. 130, pp. 94–104, Dec. 2018.

[42] Z. An, S. Li, J. Wang, Y. Xin, and K. Xu, "Generalization of deep neural network for bearing fault diagnosis under different working conditions using multiple kernel method," *Neurocomputing*, vol. 352, pp. 42–53, Aug. 2019.

[43] Y. Qi, W. You, C. Shen, X. Jiang, W. Huang, and Z. Zhu, "Hierarchical diagnosis network based on sparse deep neural networks and its application in bearing fault diagnosis," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Harbin)*, Jul. 2017, pp. 1–7.

[44] T. Han, D. Jiang, Q. Zhao, L. Wang, and K. Yin, "Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery," *Trans. Inst. Meas. Control*, vol. 40, no. 8, pp. 2681–2693, May 2018.

[45] T. Han, C. Liu, W. Yang, and D. Jiang, "Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application," *ISA Trans.*, vol. 97, pp. 269–281, Feb. 2020.

[46] T. Han, C. Liu, W. Yang, and D. Jiang, "A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults," *Knowl.-Based Syst.*, vol. 165, pp. 474–487, Feb. 2019.

[47] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 1, pp. 136–144, Jan. 2019.

[48] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, Jul. 2016.

[49] C. Tan, F. Sun, and T. Kong, "A survey on deep transfer learning," in *Proc. Int. Conf. Artif. Neural Netw.*, Cham, Switzerland: Springer, 2018, pp. 270–279.

[50] N. Segev, M. Harel, S. Mannor, K. Crammer, and R. El-Yaniv, "Learn on source, refine on target: A model transfer learning framework with random forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1811–1824, Sep. 2017.

[51] W. Ge and Y. Yu, "Borrowing treasures from the wealthy: Deep transfer learning through selective joint fine-tuning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1086–1095.

[52] 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.

[53] C.-X. Ren, D.-Q. Dai, and H. Yan, "Robust classification using $\ell_{2,1}$-norm based regression model," *Pattern Recognit.*, vol. 45, no. 7, pp. 2708–2718, Jul. 2012.

[54] M. S. Bazaraa and J. J. Goode, "A survey of various tactics for generating Lagrangian multipliers in the context of Lagrangian duality," *Eur. J. Oper. Res.*, vol. 3, no. 4, pp. 322–338, Jul. 1979.

[55] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.

[56] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, Nov. 2015, Art. no. 115002.

YI WAN (Member, IEEE) received the Ph.D. degree in materials science and engineering from the Nanjing University of Science and Technology, Nanjing, in 2008. She is currently an Associate Professor with the School of Environmental Science, Nanjing Xiaozhuang University, Nanjing. Her research interests include analysis of tribological systems, simulation planning and design, and innovation method.

TIAN-YU ZUO is currently an Undergraduate Student with the School of Mechanical and Electrical Engineering, SanJiang University, Nanjing, China. His major is mechanical engineering and automation. His current research interests include intelligent manufacturing and evolutionary algorithms.

WEI XU is currently a Senior Experimentalist with the School of Mechanical and Electrical Engineering, SanJiang University, Nanjing, China. His current research interests include intelligence manufacturing, optimal design, and virtual simulation.

XIN-MEI SHA is currently a Senior Experimentalist with the School of Mechanical and Electrical Engineering, SanJiang University, Nanjing, China. Her current research interests include intelligence manufacturing, simulation planning, and design.

• • •