A New Transfer Learning Fault Diagnosis Method Using TSC and JGSA Under Variable Condition

YUJIE YU[1], (Member, IEEE), CHUNGUANG ZHANG [1,4], (Member, IEEE),
YINGJIE LI[2], (Member, IEEE), AND YUANGANG LI[3], (Member, IEEE)
[1]School of Electronics and Information Engineering, Dalian Jiaotong University, Dalian 116028, China
[2]Spic Hubei New Energy Company Ltd., Wuhan 430077, China
[3]Business School, Dalian University of Foreign Languages, Dalian 116044, China
[4]Traction Power State Key Laboratory, Southwest Jiaotong University, Chengdu 610031, China

Corresponding author: Yuangang Li (yuangang.li@outlook.com)

ABSTRACT It is very difficult to obtain the label data of rolling bearings under the complicated and variable working conditions, which results in low diagnosis accuracy. Transfer sparse coding(TSC) is a new feature representation method, which can effectively extract features from data matrix. Joint geometric and statistical alignment (JGSA) is a domain adaptation method, which can reduce the distribution shift and geometric shift between domains. In order to make full use of the feature extraction ability of the TSC and the transfer classification ability of the JGSA, a new transfer learing fault diagnosis(TSC-JGSA) method based on combining the characteristics of the TSC and JGSA is proposed to realize the fault diagnosis of rolling bearings under variable working conditions in this paper. In the TSC-JGSA, the fast Fourier transform technology is used to transform the time-domain signals into frequency-domain amplitudes. Then the TSC is used to effectively extract the deep features from the obtained frequency-domain amplitudes in order to construct a sparse feature matrix, which is input into the JGSA in order to realize the fault diagnosis of rolling bearings. Finally, the vibration data of rolling bearings under variable working conditions is used to prove the effectiveness of the TSC-JGSA. The experiment results show that the TSC-JGSA can effecrively solve the problem of lacking label data in actual engineering by using label data in the laboratory, and obtan higher diagnosis accuracy than other compared methods. It provides a new diagnosis idea for rotating machinery.

INDEX TERMS Fault diagosis, transfer learning, domain adaptation, sparse coding, variable working conditions.

I. INTRODUCTION

Rolling bearings play an important role in large-scale rotating machinery and equipment. Once a fault occurs for rolling bearings, it will cause serious economic loss or casualties. Therefore, it is very necessary to carry out fault diagnosis for them [1]–[3]. However, the actual working conditions of rolling bearings are complex and changeable. It is of great practical significance to effectively use the labeled data of known working conditions for fault diagnosis of unlabeled data in the actual engineering under unknown working conditions.

The associate editor coordinating the review of this manuscript and approving it for publication was Baoping Cai [ID].

In recent years, many scholars have deeply studied the fault diagnosis problem under variable operating conditions [4]–[7]. The electrostatic detection method is used to study the characteristics of electrostatic signals corresponding to different fault injection degrees of rolling bearing under variable working conditions. The multi-body contact dynamic model is established to explore the multi-body contact dynamic characteristics of ball bearing under variable working conditions. The parameter optimized variational mode decomposition (POVMD) and envelope order spectrum are used to extract fault feature information. The vibration signal of rolling bearings is analyzed and processed by fast spectral correlation to obtain the eigenvector, which is input into PSO-SVM for state recognition.

However, the traditional fault dianosis methods have the limitations of complicated analysis process and poor generalization ability under variable working conditions. Transfer learning has attracted extensive attention [8], [9]. The feature extraction method of singular value decomposition and autocorrelation matrix is combined with the transfer learning TrAdaBoost algorithm to diagnose the motor fault [10]–[14]. A new fault identification method based on the combining long and short term memory network (LSTM) and transfer learning (TL) is proposed. A semi-supervised integrated learning tool (SSIT) based on transfer learning is proposed for engine bearing fault prediction to solve the problems of insufficient prediction accuracy and over-fitting [15]–[17]. A bearing fault diagnosis model based on the transfer learning is composed of a stack sparse automatic encoder and a flexible maximum function regression.

Domain adaption is an important branch of transfer learning. It is an effective method to solve the problem that the source domain and the target domain have different data distribution but the same task. The subspace described by the eigenvector is used to represent the source domain and the target domain [18]. A mapping function is learned to find a domain adaptation solution. The mapping function aligns the source subspace with the target subspace. The subspace mapping method is extended. The distribution and subspace basis of source domain and target domain are aligned simultaneously [19]. A kernel-based method is proposed to utilize the inherent low-dimensional structure of data sets [20]. The geodesic flow kernel model describes the changes in geometric and statistical characteristics from the source domain to the target domain by integrating infinite subspaces [21]. A new dimension reduction method, transfer component analysis, is proposed to find a good cross-domain feature representation [22]. The joint distribution adaptation method adapts marginal distribution and conditional distribution in the process of dimension reduction according to certain rules. It constructs a new feature representation. Feature matching and sample re-weighting are established in a unified optimization problem. A transfer joint matching method is proposed [23]. A joint geometric and statistical alignment (JGSA) algorithm is proposed to reduce domain shift both statistically and geometrically [24].

Recently, except for the mentioned feature-based transfer learning methods, domain adaptation is realized through many new ideas. Meanwhile, more and more scholars focus on neural networks and adversarial learning. An algorithm named structured domain adaptation (SDA) is proposed to seek a discriminate subspace shared by two domains where the well-learned knowledge of the source domain can be transferred to the target domain [25]. Samples from both domains are combined together to reveal more shared information across two domains. A previously unexplored instance of the general framework is proposed which combines discriminative modeling, untied weight sharing [26]. A novel deep learning framework that can exploit labeled source data and unlabeled target data

to learn informative hash codes is proposed to accurately classify unseen target data [27]. A weighted MMD model is proposed by introducing an auxiliary weight for each class in the source domain [28]. A novel approach named Learning Distribution-Matched Landmarks (LDML) is proposed [29]. LDML reveals the latent factors by learning a domain-invariant subspace where the two domains are well aligned at both feature level and sample level. A two-stream architecture is introduced where one operates in the source domain and the other in the target domain [30]. In contrast to other approaches, the weights in corresponding layers are related but not shared. A new unsupervised domain adaptation approach called Collaborative and Adversarial Network (CAN) is proposed through domain-collaborative and domain-adversarial training of neural networks [31]. A new domain adaptation method called Domain-Symmetric Networks (SymNets) is proposed [32]. Transferrable Prototypical Networks (TPN) is presented for adaptation such that the prototypes for each class in the two domains are close in the embedding space and the score distributions predicted by prototypes separately on source and target data are similar [33]. Contrastive Adaptation Network (CAN) is proposed to optimize a new metric which explicitly models the intra-class domain discrepancy and the inter-class domain discrepancy [34]. A dynamic Bayesian network (DBN)-based fault diagnosis methodology in the presence of TF and IF for electronic systems is proposed [35]. An OOBN-based real-time fault diagnosis methodology is proposed [36]. A hybrid physics-model-based and data-driven remaining useful life (RUL) estimation methodology of structure systems by using dynamic Bayesian networks (DBNs) is proposed [37]. A personalized diagnosis method to detect faults in gears using numerical simulation and extreme learning machine is proposed [38]. A personalized diagnosis method to detect faults in a bearing based on acceleration sensors and an FEM simulation driving support vector machine is proposed [39]. A new unsupervised domain adaptation method named domain-adversarial residual-transfer learning of deep neural networks is proposed to tackle cross-domain image classification tasks [40]. A domain adaptation method for machinery fault diagnostics based on deep learning is proposed to address the fault diagnostic tasks with data from different places of machines [41], [42]. A domain adaptation diagnostic model based improved deep neural network is proposed, which diagnoses early gear pitting faults under multiple working conditions [43], [44]. In addition, some new algorithms are also proposed, which can optimize the diagnosis models to improve the classification performance [45]–[52].

It can be seen from the previous work that the deep learning method is introduced to solve the transfer problem, which can achieve ideal results by better feature extraction effect. Transfer sparse coding (TSC) is proposed to construct robust sparse representations for classifying cross-distribution images accurately [53]. It is a new feature representation method, which can effectively extract

features from data matrix. The JGSA is a domain adaptation method, which can reduce the distribution shift and geometric shift between domains by using shared features and domain-specific features. It can effectively and accurately realize the classification. In order to make full use of the feature extraction ability of the TSC and the transfer classification ability of the JGSA, the TSC and JGSA are conbined and introduced into fault diagnosis to propose a new transfer learning fault diagnosis (TSC-JGSA) method for rolling bearings under variable working conditions. In the proposed TSC-JGSA method, the time-domain vibration signals under known and unknown working conditions are preprocessed by fast Fourier transform to obtain the frequency-domain amplitudes. Then the TSC is used to extract features from frequency-domain amplitudes to construct the sparse feature matrix, which is input into the JGSA model to realize a new fault diagnosis method. Finally, the vibration signals of rolling bearings under variable working conditions are used to prove the effectiveness of the TSC-JGSA method.

The main contributions of this paper are summarized as follows:

• A new transfer learning fault diagnosis method based on creatively combining sparse coding method with the domain adaptation algorithm is proposed to realize the fault classification.

• The motivation to combine TSC and JGSA is to make full use of the feature extraction ability of the TSC and the transfer classification ability of the JGSA to realize the fault diagnosis of rolling bearings under variable working conditions.

• The performance of the TSC-JGSA method is has been extensively investigated by the vibration signals of rolling bearings under variable working conditions.

• The TSC-JGSA method can effecrively solve the problem of lacking label data in actual engineering by using label data in the laboratory.

II. BASIC METHODS
A. TSC

Sparse coding is an effective method in feature extraction, which can realize adaptive feature extraction [53], [54]. Given a data matrix $X = [x_1, \ldots, x_n] \in R^{m \times n}$, it means that n data points are sampled from the m-dimension feature space, and the dictionary matrix is expressed as $B = [b_1, \ldots, b_k] \in R^{m \times k}$, where each column vector b_i represents a basis vector in the dictionary, and the coding matrix is expressed as $S = [s_1, \ldots, s_n] \in R^{k \times n}$, where each column vector s_i is a sparse representation of a data point x_i. The specific formula is given as follow.

$$min_{B,S}||X - BS||_F^2 + \lambda \sum_{i=1}^{n} |s_i|$$
$$\text{s.t.} ||b_i||^2 \le c, \quad \forall i = 1, \ldots, k \qquad (1)$$

where λ is an adjustable regularization parameter, which is used to determine the sparsity of coding and the approximation of input data.

Sparse coding learns a set of dictionary basis vectors and the coding matrix from given training samples by

loop iteration. Then it uses the optimized method to solve the coding matrix of the test samples. Through this process, the raw data is reconstructed into a new feature representation using dictionary basis vectors. The coding matrix S is a sparse matrix, which can greatly improve the calculation speed and save storage space. It is the advantage of sparse matrix.

In order to take into account the potential intrinsic geometry of the input data, a graph regularized sparse coding (GraphSC) method is proposed [54]. This method obtains data geometric information by learning the sparse representation that explicitly considers the local manifold structure of the data. The formula is given as follow.

$$min_{B,S}||X - BS||_F^2 + \gamma tr\left(SLS^T\right) + \lambda \sum_{i=1}^{n} |s_i|$$
$$\text{s.t.} ||b_i||^2 \le c, \quad \forall i = 1, \ldots, k \qquad (2)$$

where γ is a graph regularization parameter, which is used to determine the weight between sparse coding and geometry preservation.

However, when the labeled and unlabeled data are sampled from different distributions, they may be quantized into different feature words of the codebook and encoded with different representations, which may seriously degrade the classification performance. To improve the learning ability of sparse coding, we minimize the distribution difference between the labeled data and unlabeled data, and apply this criterion to the objective function of sparse coding [45]. The empirical maximum mean difference (MMD) is used as a nonparametric distance measure to compare different distributions, and the objective function of TSC is obtained as follow.

$$min_{B,S}||X - BS||_F^2 + tr\left(S\left(\mu M + \gamma L\right)S^T\right) + \lambda \sum_{i=1}^{n} |s_i|$$
$$\text{s.t.} ||b_i||^2 \le c, \quad \forall i = 1, \ldots, k \qquad (3)$$

where $\mu > 0$ is the MMD regularization parameter, which is used to determine the weight between GraphSC and distribution matching.

B. JGSA

In order to reduce the shift between domains both geometrically and statistically, a domain adaptation method, referred as joint geometric and statistical alignment, namely JGSA is proposed by using both shared features and domain-specific features. The definition of the JGSA is described as follows. Source domain data is represented as X_s, extracted from distribution $P_s(X_s)$, and target domain data is represented as X_t, obtained from distribution $P_t(X_t)$, where n_s and n_t are number of samples in source domain and target domain, respectively. We assume that the feature space and label space of the two domains are the same as $X_s = X_t$ and $Y_s = Y_t$. Due to data set offset, there is $P_s(X_s) \ne P_t(X_t)$.

The JGSA obtains a new representation of each domain by finding two coupling projections (A for the source domain and B for the target domain), mapping the source domain data and the target domain data to their respective subspaces.

After projection, 1) Maximize the variance of the target domain data to maintain the target domain data attributes. 2) The discrimination information of the source domain data is retained to transfer the label information effectively. 3) Minimize both the marginal distribution and conditional distribution divergences between the source and target domains to statistically reduce domain shift. 4) The divergence between the two projections is limited as small as possible, and the shift is reduced geometrically.

Based on the above four points, the objective function of the JGSA method is obtained as follow.

$$max_{A,B} \frac{Tr([A^T\ B^T][\begin{smallmatrix} \beta S_b & 0 \\ 0 & \mu S_t \end{smallmatrix}][\begin{smallmatrix} A \\ B \end{smallmatrix}])}{Tr([A^T\ B^T][\begin{smallmatrix} M_s + \lambda I + \beta S_w & M_{st} - \lambda I \\ M_{ts} - \lambda I & M_t + (\lambda + \mu)I \end{smallmatrix}][\begin{smallmatrix} A \\ B \end{smallmatrix}])} \quad (4)$$

The goal is to find two coupled projections A and B by solving the optimization function, where $I \in R^{d \times d}$ is the identity matrix, μ is the target domain scatter matrix coefficient, β is the within class scatter matrix and between class scatter matrix coefficient, and λ is the subspace shift coefficient.

III. A TRANSFER LEARNING FAULT DIAGNOSIS METHOD
A. THE IDEA OF THE FAULT DIAGNOSIS
In the case of variable working conditions, there is a large difference in the statistical distribution between the data under known and unknown working conditions, but the feature space and the label space are the same. The potential common features from the known working conditions is learned. At the same time, the unique features of the unknown working conditions is retained to improve the generalization ability of the model. The vibration signal is transformed from time-domain to frequency-domain, so as to obtain fault information, which is more conducive to identification and matching. This transformation is also of practical significance for reducing the computational complexity of subsequent algorithms. The data collection environment is not a strict laboratory environment, it is closer to the actual engineering conditions. So the fault data often contains more redundant information and is accompanied by noise, which needs to be further extracted. The sparse coding is superior in feature extraction. TSC is a new feature representation method, which can achieve the desired feature extraction effect. It can find and extract effective features from the data, which also eliminates the influence of noise. It is a powerful tool to extract useful features from data, thus can more accurately and concisely express data features. The domain adaptation method is an effective method to solve the problem that the data distribution in the source domain is different from that in the target domain, but the tasks are the same. The JGSA method can reduce the distribution shift and geometrical shift between domains by simultaneously using the shared features and domain-specific features between source domain data and target domain data to achieve the purpose of transfer learning. In order to make full use of the feature extraction

ability of the TSC and the transfer classification ability of the JGSA, a new transfer learning fault diagnosis (TSC-JGSA) method based on creatively combining TSC with JGSA is proposed to realize the fault diagnosis of rolling bearings under variable working conditions. The bearing vibration data of three different rotating speeds and ten fault types under known and unknown working conditions are used to verify the effectiveness of the TSC-JGSA method.

B. FAULT DIAGNOSIS MODEL
The flow of the TSC-JGSA method for the fault diagnosis of rolling bearings under variable condition is shown in Figure 1.
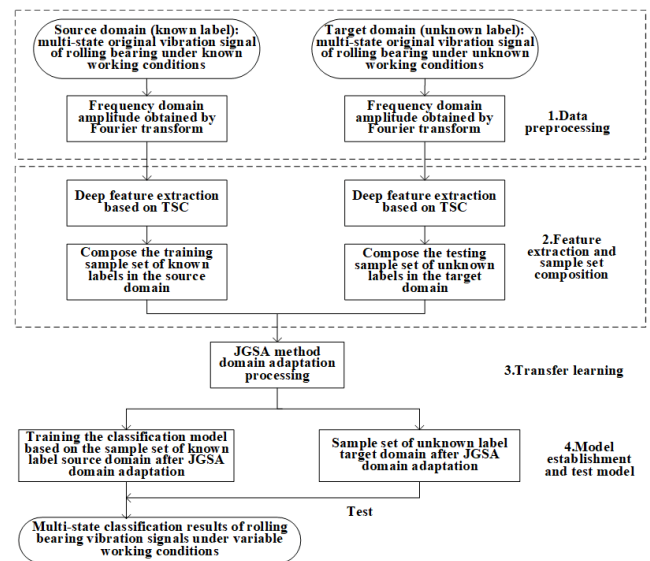

FIGURE 1. The flow of TSC-JGSA.

C. THE STEPS OF THE TSC-JGSA
In this paper, the advantages of sparse coding are used to extract deep features. The variable working condition is solved by combining the characteristics of the TSC and JGSA, the differences of samples between different working conditions are reduced to realize fault diagnosis with various types for rolling bearings at different rotating speeds. The specific steps of the TSC-JGSA are described as follows.

Step 1. Data preprocessing

The multi-state time-domain vibration signals of rolling bearings under known and unknown working conditions are transformed into frequency-domain by using fast Fourier transform in order to obtain corresponding frequency-domain amplitudes.

Step 2. Feature extraction

The TSC is used to extract the deep features of frequency domain amplitudes of the vibration signals.

Step 3. Construct feature matrix

The obtained features are used to construct the feature matrix, which is composed of the training samples(source domain) and the test sample(target domain).

Step 4. Domain adaptation processing

The JGSA is used to perform domain adaptation processing on the training samples(source domain) and the test samples(target domain), which statistically reduce the marginal distribution domain shift and the conditional distribution domain shift between different domains, geometrically reduces the subspace domain shift after projection, and improves the distribution similarity of the samples between domains.

Step 5. Model establishment

The k-NearestNeighbor (KNN) classification model is trained by the known label samples in the source domain after JGSA domain adaptation, and the unknown label samples of target domain are used to test model after domain adaptation. Through continuous iterative calculation, multi-state classification results of bearings under variable working conditions are obtained on the target domain samples.

IV. VALIDATION AND ANALYSIS
A. DATA DESCRIPTION AND PARAMETER SETTING

In this experiment, the obtained experimental data come from the experiment platform of QPZZ-IIrotating machinery. The experiment platform of QPZZ-IIrotating machinery is shown in Figure 2. The experiment platform of QPZZ-II rotating machinery consists of a variable-speed drive machine with a power of 0.75KW, bearings, gearboxes, shafts, eccentric turntables, governor, and so on.
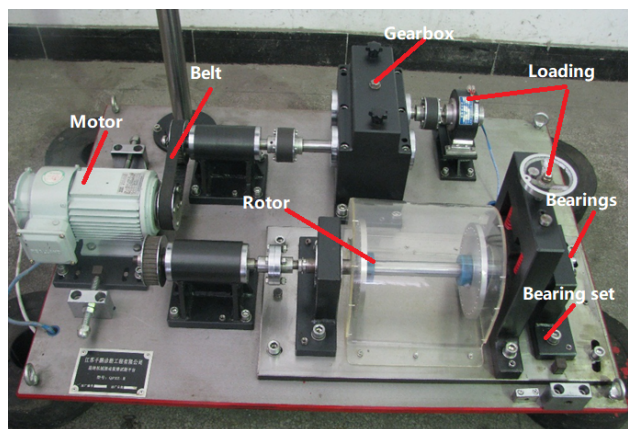

FIGURE 2. The experiment platform of QPZZ-IIrotating machinery.

The inner ring, outer ring and rolling elements of the bearing are processed with single-point damage. The eccentric turntable simulates the unbalanced working condition of the rotor by placing weights. The rolling bearing is a cylindrical roller rolling bearing of type N205. By simulating the imbalance between the fault bearing and the shafting, the acceleration vibration signals at different rotating speeds and different fault positions are collected. The fault positions are at the inner ring, outer ring and rolling element of the rolling bearings. The vibration signal is collected by vibration acceleration sensor and data acquisition card, and the sampling frequency is 12kHZ. The bearing fault

also contains data with a sampling frequency of 48kHZ. Record the vibration acceleration signal data under the motor speed conditions of 1000rpm, 1250rpm and 1500rpm, respectively.

At each speed, the vibration signal data of 10 states are divided into 5 single faults of normal, inner ring, outer ring, rolling element and rotor imbalance and 5 coupling faults. The data is intercepted with 1024 length. A total of 10 classifications were performed on the input data. The states and label descriptions of the 10 classifications are shown in Table 1. The environment of this experiment is the computer processor Intel (R) Core (TM) i5-7400 CPU @ 3.00GHz, memory 8GB, MATLAB 2018b.

TABLE 1. Data descriptions.

Fault types	Name	Labels
Inner ring	IR	1
Outer ring	OR	2
Rolling element	RE	3
Outer ring and rolling element	ORE	4
Rotor unbalance	RUB	5
Rotor unbalance and inner ring	RUI	6
Rotor unbalance and outer ring	RUO	7
Rotor unbalance and rolling element	ROR	8
Rotor unbalance, outer ring and rolling element	RUOR	9
Normal state	NORL	0

10 kinds of multi-state samples of rolling bearings under different working conditions are set as follows.

(1) Set the working condition A as 1000r/min.
(2) Set the working condition B as 1250r/min.
(3) Set the working condition C as 1500r/min.

TABLE 2. Sample set composition of rolling bearing.

Different working condition	Source domain	Target domain	Number of source domain samples	Number of source domain samples
A/B working condition	A	B		
B/A working condition	B	A		
B/C working condition	B	C		
C/B working condition	C	B	2000	1800
A/C working condition	A	C		
C/A working condition	C	A		

The specific descriptions of the vibration signal samples are shown in Table 2. In Table 2, the A/B working condition indicates that the working condition A feature sample set in multiple states is used as the source domain, that is, the training feature sample set, and the working condition B feature sample set is used as the target domain, that is, the test feature sample set. Select 2000 samples in the source domain for training and 1800 samples in the target domain for testing under each working condition. The reason for this choice is that the fault diagnosis accuracy can achieve better results when the training data is slightly more than the test data.

The time-domain data samples of rolling bearing vibration signals under the above-mentioned different working conditions are transformed into the frequency-domain data samples by using fast Fourier transform. At this time, the dimension of the frequency domain amplitudes is half of the original time-domain dimension, namely 513 dimensions. Input the transformed frequency- domain amplitudes of all samples as the characteristic values into TSC.

In this paper, the **values of parameters** are set according to the suggestions given by the original authors of the two methods. It is generally believed that as long as they are within the given range, they can achieve good results. The subspace basis vector of JGSA is set as 30, and the kernel type is set to primeval, which represents the results of the JGSA in the original data space. In JGSA, the target domain scatter matrix coefficient μ is 1, the within class scatter matrix and between class scatter matrix coefficient β is 0.1, and the subspace shift coefficient λ is 1. The number of PCA basis vectors in TSC is set as 64, the dictionary dimension is k, the number of dictionary basis vectors is 128, the MMD regularization parameter is 1e6, the graph regularization parameter is 1, the sparsity penalty parameter is 0.1, and the number of TSC iterations is 10.

B. FEATURE EXTRACTION

The TSC can adaptively extract the features from the data under the unknown data label condition, which can avoid the empirical dependence of signal processing methods and improve the feature extraction efficiency. Therefore, in this paper, the features of distinguishable multi-state vibration signals obtained by TSC are used as the input of the domain adaptation model of JGSA. The flow of deep feature extraction of rolling bearing vibration signals using the TSC is shown in Figure 3.

The TSC is adopted to extract the vibration signal feature of rolling bearing. The specific steps are described as follows.

Step 1. Select the original multi-state data of rolling bearings. Set the number of vibration signal data points in each state and the length of each sample.

Step 2. According to the equation(1), the time-domain data is transformed into the frequency-domain data by using fast Fourier transform. The obtained frequency-domain data is used as the input data of TSC.

Step 3. Initialize the network structure parameters of the TSC. The frequency-domain data of the vibration signal is used as training samples. The TSC is trained in order to obtain the optimal outputs B and S when the objective function is minimized.

Step 4. The output result of the data is the sparse matrix S, which is the deep features extracted by TSC.

C. EXPERIMENT RESULTS

After the frequency-domain amplitudes of rolling bearing vibration signals is input ito the TSC to obtain deep feature samples. The JGSA is introduced to process the deep feature samples under different working conditions to achieve the
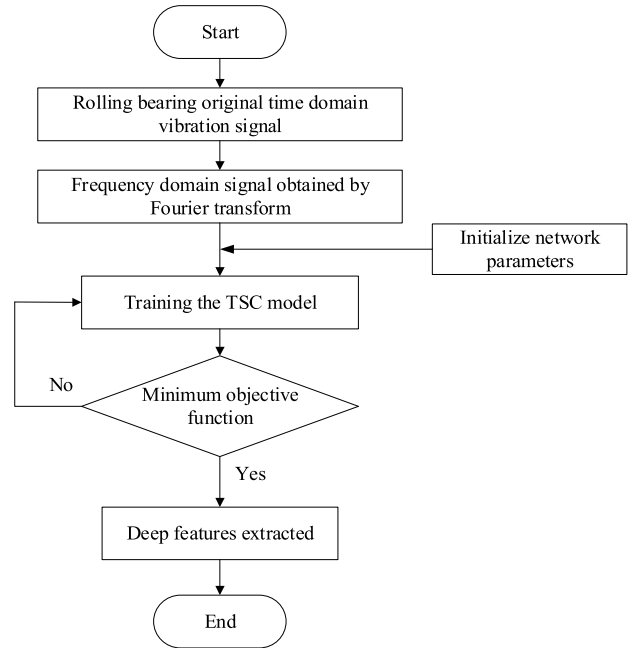

FIGURE 3. The flow of deep feature extraction.

TABLE 3. Diagnosis accuracies under different working conditions.

Different schemes	A/B (%)	B/A (%)	B/C (%)	C/B (%)	A/C (%)	C/A (%)	Average accuracy(%)
TSC-KNN	79.28	84.90	85.21	89.50	72.47	76.71	81.35
TSC-PCA	78.83	84.42	84.75	87.71	70.99	77.52	80.70
TSC-GFK	85.49	78.93	87.18	83.86	79.29	73.14	81.32
TSC-TJM	96.03	96.56	96.18	91.82	92.94	91.47	94.17
TSC-JDA	95.65	97.17	96.20	89.74	92.87	91.36	93.83
TSC-JGSA	**99.12**	**98.28**	**97.33**	**98.00**	**91.68**	**96.10**	**96.75**

increasing purpose within class compactness and between class discrimination. In order to reflect the advantages of JGSA, the non-transfer learning scheme of TSC-KNN was selected to compare with the schemes of principal components analysis (PCA), GFK, TJM and JDA. The fault diagnosis accuracies of rolling bearings using different schemes under different working conditions are shown in Table 3 and Figure 4.

As can be seen from Figure 4 and Table 3 that compared with the other five schemes, JGSA has the highest fault diagnosis accuracy under five working conditions except for the working condition A/C. Its accuracy is slightly lower than the two schemes of TJM and JDA under the working condition A/C. By the comprehensive comparison, JGSA has the highest average accuracy of fault diagnosis, which is higher than the other five comparison schemes. It can be seen from the analysis results that the deep features obtained by TSC are the same. Other methods do not well retain the attributes of the original data when the distribution shift is reduced. It fail to consider the correlation between the statistical
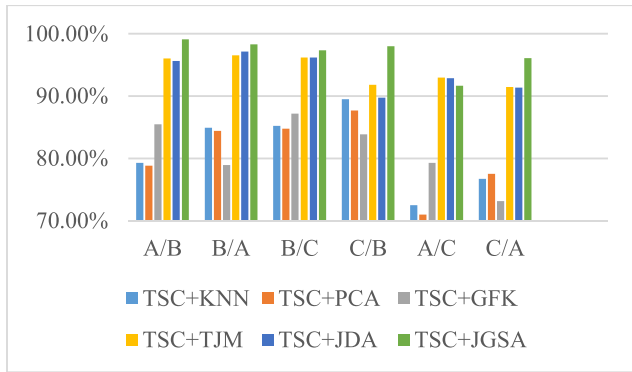
FIGURE 4. Diagnosis accuracies under different working conditions.

distribution and the geometric subspace. However, from the perspective of feature transfer, the JGSA comprehensively considers the geometric and statistical characteristics of the source domain data and target domain data, and uses shared and domain-specific features to reduce domain shift. The samples under different working conditions are well aligned in the statistical distribution and geometric subspace, so that the proposed deep transfer learning has the highest accuracy of fault diagnosis in multi-state vibration signals of rolling bearing under variable working conditions.

Due to the randomness of the TSC, in order to prove that the experimental results are not accidental but universal, and also to further explore the robustness of the proposed scheme, each experiment method is repeated 10 times, and the box plots of six experimental schemes under six working conditions according to the results of 10 experiments are shown in Figure 5∼Figure 10.
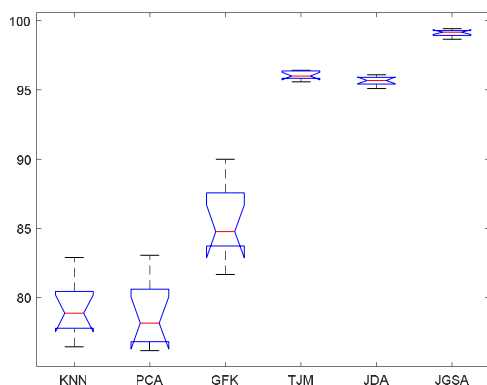

FIGURE 5. Diagnosis accuracy under A/B working condition.

As can be seen from Figure 5 to Figure 8, under the four working conditions, the lowest accuracy rate of the TSC-JGSA in 10 experiments is still higher than those of other comparative methods. Therefore, the experimental results are less discrete, which shows that the TSC-JGSA does not only have the highest accuracy in the experimental results under the four conditions, but also shows good robustness.
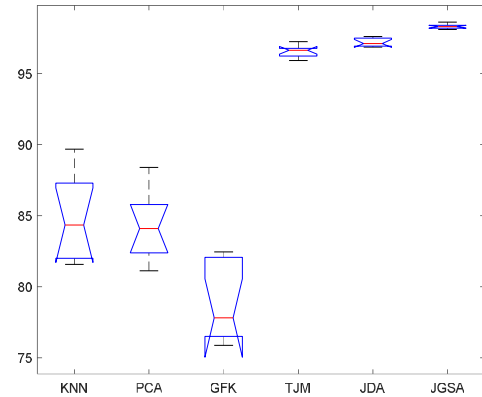

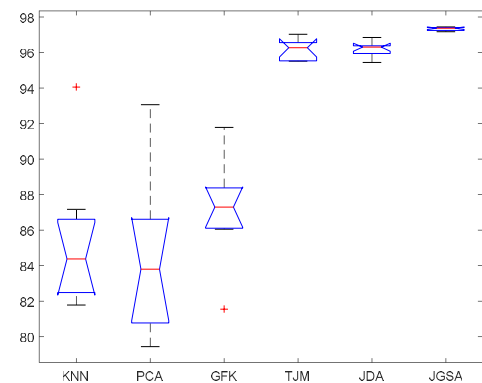**FIGURE 6.** Diagnosis accuracy under B/A working condition.

FIGURE 7. Diagnosis accuracy under B/C working condition.
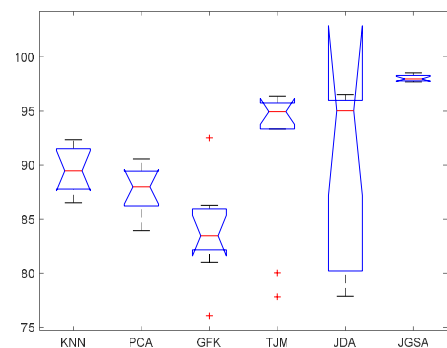

FIGURE 8. Diagnosis accuracy under C/B working condition.

It can be seen from Figure 9 that under the working condition A/C, the JGSA shows a large dispersion, including the lowest accuracy rate than the TJM method, and the median accuracy rate than the other two methods, namely TJM and JDA. It shows that when the working conditions are greatly different, that is, when the two rotating speeds are different, the JGSA can not show strong robustness. Especially the high speed data learns from the low speed data. The reason may be that the data of the two speeds are quite different, and the features can not be well aligned to reduce the distribution domain shift and the subspace domain shift. This is also a limitation of the scheme. How to enhance the robustness of
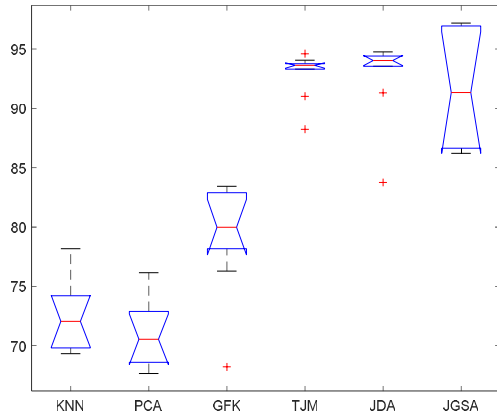
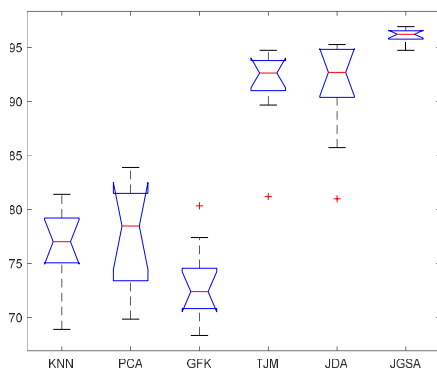**FIGURE 9.** Diagnosis accuracy under A/C working condition.

FIGURE 10. Diagnosis accuracy under C/A working condition in different schemes.

the JGSA under two different working conditions is the focus of the next research.

It can be seen from Figure 10 that the accuracy of the JGSA is lower than that of the JDA in the 10 experiments. It is still the robustness problem of the proposed method when the reflected working conditions in Figure 9 are quite difficult. However, the dispersion of the experiment results and the median accuracy rate have been greatly improved by comparing the results of the previous working condition, which is still higher than the other five comparative experimental methods.

V. CONCLUSION AND PROSPECTS

In this paper, a new transfer learning fault diagnosis method based on TSC and JGSA for rolling bearings under variable working conditions is proposed. The TSC is used to extract the vibration signal features of rolling bearings under variable working conditions, so as to obtain features that can better represent the state of rolling bearing. The JGSA domain adaptation method is introduced to reduce the distribution domain shift and subspace domain shift simultaneously, and the samples under different working conditions are well aligned statistically and geometrically to reduce sample differences between different working conditions. Compared with other domain adaptation methods and non-transfer

learning methods, the experimental results show that the proposed TSC-JGSA method has higher accuracy and stronger robustness in rolling bearing state classification under variable speed.

In the next step, we will further study the domain adaptation method in the transfer learning to improve the robustness of the JGSA in the fault diagnosis under variable working conditions, so as to better diagnose rolling bearing fault diagnosis under variable speed.

REFERENCES

[1] W. Deng, H. Liu, J. Xu, H. Zhao, and Y. Song, "An improved quantum-inspired differential evolution algorithm for deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 7319–7327, Oct. 2020.

[2] H. Zhao, J. Zheng, J. Xu, and W. Deng, "Fault diagnosis method based on principal component analysis and broad learning system," *IEEE Access*, vol. 7, pp. 99263–99272, 2019.

[3] S. Haidong, C. Junsheng, J. Hongkai, Y. Yu, and W. Zhantao, "Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing," *Knowl.-Based Syst.*, vol. 188, Jan. 2020, Art. no. 105022.

[4] W. Deng, R. Yao, H. Zhao, X. Yang, and G. Li, "A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm," *Soft Comput.*, vol. 23, no. 7, pp. 2445–2462, Apr. 2019.

[5] Z. He, H. Shao, X. Zhang, J. Cheng, and Y. Yang, "Improved deep transfer auto-encoder for fault diagnosis of gearbox under variable working conditions with small training samples," *IEEE Access*, vol. 7, pp. 115368–115377, 2019.

[6] H. Zhao, S. Zuo, M. Hou, W. Liu, L. Yu, X. Yang, and W. Deng, "A novel adaptive signal processing method based on enhanced empirical wavelet transform technology," *Sensors*, vol. 18, no. 10, p. 3323, Oct. 2018.

[7] Y. Hu and C. Zhao, "Fault diagnosis with dual cointegration analysis of common and specific nonstationary fault variations," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 1, pp. 237–247, Jan. 2020.

[8] H. Zhao, D. Li, W. Deng, and X. Yang, "Research on vibration suppression method of alternating current motor based on fractional order control strategy," *Proc. Inst. Mech. Eng., Part E, J. Process Mech. Eng.*, vol. 231, no. 4, pp. 786–799, Aug. 2017.

[9] W. Qian, S. Li, and J. Wang, "A new transfer learning method and its application on rotating machine fault diagnosis under variant working conditions," *IEEE Access*, vol. 6, pp. 69907–69917, 2018.

[10] H. Zhao, H. Liu, J. Xu, and W. Deng, "Performance prediction using high-order differential mathematical morphology gradient spectrum entropy and extreme learning machine," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4165–4172, Jul. 2020.

[11] G. Xu, M. Liu, Z. Jiang, W. Shen, and C. Huang, "Online fault diagnosis method based on transfer convolutional neural networks," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 509–520, Feb. 2020.

[12] H. Zhao, J. Zheng, W. Deng, and Y. Song, "Semi-supervised broad learning system based on manifold regularization and broad network," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 3, pp. 983–994, Mar. 2020.

[13] 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 CNN and thermal images," *IEEE Trans. Ind. Informat.*, early access, Jun. 30, 2020, doi: 10.1109/TII.2020.3005965.

[14] 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.

[15] X. Li, W. Zhang, Q. Ding, and X. Li, "Diagnosing rotating machines with weakly supervised data using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1688–1697, Mar. 2020.

[16] X. Wang, X. Liu, and Y. Li, "An incremental model transfer method for complex process fault diagnosis," *IEEE/CAA J. Automatica Sinica*, vol. 6, no. 5, pp. 1268–1280, Sep. 2019.

[17] Z. Chen, K. Gryllias, and W. Li, "Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 339–349, Jan. 2020.

[18] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2960–2967.

[19] B. Sun and K. Saenko, "Subspace distribution alignment for unsupervised domain adaptation," in *Proc. Brit. Mach. Vis. Conf.*, 2015, p. 24, doi: 10.5244/C.29.24.

[20] 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.

[21] 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.

[22] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2200–2207.

[23] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1410–1417.

[24] J. Zhang, W. Li, and P. Ogunbona, "Joint geometrical and statistical alignment for visual domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR).* Jul. 2017, pp. 5150–5158.

[25] J. Li, Y. Wu, and K. Lu, "Structured domain adaptation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 8, pp. 1700–1713, Aug. 2017.

[26] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2962–2971.

[27] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, "Deep hashing network for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5385–5394.

[28] H. Yan, Y. Ding, P. Li, Q. Wang, Y. Xu, and W. Zuo, "Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 945–954.

[29] M. Jing, J. Li, J. Zhao, and K. Lu, "Learning distribution-matched landmarks for unsupervised domain adaptation," in *Database Systems for Advanced Applications.* Cham, Switzerland: Springer, 2018, pp. 491–508.

[30] A. Rozantsev, M. Salzmann, and P. Fua, "Beyond sharing weights for deep domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, pp. 801–814, Apr. 2019.

[31] W. Zhang, W. Ouyang, W. Li, and D. Xu, "Collaborative and adversarial network for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3801–3809.

[32] Y. Zhang, H. Tang, K. Jia, and M. Tan, "Domain-symmetric networks for adversarial domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5031–5040.

[33] Y. Pan, T. Yao, Y. Li, Y. Wang, C.-W. Ngo, and T. Mei, "Transferrable prototypical networks for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2234–2242.

[34] G. Kang, L. Jiang, Y. Yang, and A. G. Hauptmann, "Contrastive adaptation network for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4888–4897.

[35] B. Cai, Y. Liu, and M. Xie, "A dynamic-Bayesian-network-based fault diagnosis methodology considering transient and intermittent faults," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 1, pp. 276–285, Jan. 2017.

[36] B. Cai, H. Liu, and M. Xie, "A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks," *Mech. Syst. Signal Process.*, vol. 80, pp. 31–44, Dec. 2016.

[37] B. Cai, X. Shao, Y. Liu, X. Kong, H. Wang, H. Xu, and W. Ge, "Remaining useful life estimation of structure systems under the influence of multiple causes: Subsea pipelines as a case study," *IEEE Trans. Ind. Electron.*, vol. 67, no. 7, pp. 5737–5747, Jul. 2020.

[38] X. Liu, H. Huang, and J. Xiang, "A personalized diagnosis method to detect faults in gears using numerical simulation and extreme learning machine," *Knowl.-Based Syst.*, vol. 195, May 2020, Art. no. 105653.

[39] X. Liu, H. Huang, and J. Xiang, "A personalized diagnosis method to detect faults in a bearing based on acceleration sensors and an FEM simulation driving support vector machine," *Sensors*, vol. 20, no. 2, p. 420, Jan. 2020.

[40] X. Fang, H. Bai, Z. Guo, B. Shen, S. Hoi, and Z. Xu, "DART: Domain-adversarial residual-transfer networks for unsupervised cross-domain image classification," *Neural Netw.*, vol. 127, pp. 182–192, Jul. 2020, doi: 10.1016/j.neunet.2020.03.025.

[41] X. Li, W. Zhang, N.-X. Xu, and Q. Ding, "Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places," *IEEE Trans. Ind. Electron.*, vol. 67, no. 8, pp. 6785–6794, Aug. 2020.

[42] J. Li, X. Li, D. He, and Y. Qu, "A domain adaptation model for early gear pitting fault diagnosis based on deep transfer learning network," *Proc. Inst. Mech. Eng., Part O, J. Risk Rel.*, vol. 234, no. 1, pp. 168–182, 2020.

[43] Y. Liu, X. Wang, Z. Zhai, R. Chen, B. Zhang, and Y. Jiang, "Timely daily activity recognition from headmost sensor events," *ISA Trans.*, vol. 94, pp. 379–390, Nov. 2019.

[44] B. Quanz, J. Huan, and M. Mishra, "Knowledge transfer with low-quality data: A feature extraction issue," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 10, pp. 1789–1802, Oct. 2012.

[45] W. Deng, W. Li, and X.-H. Yang, "A novel hybrid optimization algorithm of computational intelligence techniques for highway passenger volume prediction," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4198–4205, Apr. 2011.

[46] W. Deng, J. J. Xu, H. M. Zhao, and Y. J. Song, "A novel gate resource allocation method using improved PSO-based QEA," *IEEE Trans. Intell. Transp. Syst.*, to be published, doi: 10.1109/TITS.2020.3025796.

[47] W. Deng, J. Xu, Y. Song, and H. Zhao, "Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem," *Appl. Soft Comput.*, early access, Sep. 15, 2020, doi: 10.1016/j.asoc.2020.106724.

[48] Y. Liu, Y. Mu, K. Chen, Y. Li, and J. Guo, "Daily activity feature selection in smart homes based on Pearson correlation coefficient," *Neural Process. Lett.*, vol. 51, no. 2, pp. 1771–1787, Apr. 2020.

[49] W. Deng, J. Xu, and H. Zhao, "An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem," *IEEE Access*, vol. 7, pp. 20281–20292, 2019.

[50] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2019.

[51] R. Chen, S.-K. Guo, X.-Z. Wang, and T.-L. Zhang, "Fusion of multi-RSMOTE with fuzzy integral to classify bug reports with an imbalanced distribution," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 12, pp. 2406–2420, Dec. 2019.

[52] W. Deng, J. Xu, Y. Song, and H. Zhao, "An effective improved co-evolution ant colony optimization algorithm with multi-strategies and its application," *Int. J. Bio-Inspired Comput.*, pp. 1–10, 2020.

[53] M. Long, G. Ding, J. Wang, and J. Sun, Y. Guo, and S. Y. Philip, "Transfer sparse coding for robust image representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 407–414.

[54] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai, "Graph regularized sparse coding for image representation," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1327–1336, May 2011.

YUJIE YU (Member, IEEE) received the B.S. degree in electrical engineering and automation from the North University of China, Taiyuan, China, in 2018. She is currently pursuing the master's degree with Dalian Jiaotong University, Dalian, China. Her research interests include deep learning and fault diagnosis.

CHUNGUANG ZHANG (Member, IEEE) received the B.S. degree in electrical technology and the M.S. degree in traffic information control engineering from Dalian Jiaotong University, Dalian, China, in 2001 and 2005, respectively. Since 2014, he has been an Associate Professor with the School of Electronics and Information Engineering, Dalian Jiaotong University. His research interests include signal processing and fault diagnosis.

YINGJIE LI (Member, IEEE) received the bachelor's degree in power engineering from North China Electric Power University, Beijing, in 1997. He was a Senior Engineer. Since 1997, he has been the Director of the Production Technology Department, State Power Investment Group, Pingdingshan Thermal Power Company Ltd., and a Deputy General Manager with SPIC Hubei New Energy Company Ltd. His research interests include on-line fault diagnosis, early warning of power generation equipment, and power system analysis.

YUANGANG LI (Member, IEEE) received the B.S. degree in computer science and technology from the Dalian University of Technology, China, in 1997, the M.B.A. degree from the Dongbei University of Finance and Economics, China, in 2006, and the Ph.D. degree in management science and engineering from the Shanghai University of Finance and Economics, China, in 2018. He is currently a Lecturer with the Business School, Dalian University of Foreign Languages, Dalian China. His research interests include complex system modeling, signal processing, data mining, and data analysis.