Deep Adversarial Domain Adaptation Model for Bearing Fault Diagnosis

Zhao-Hua Liu, *Member*, *IEEE*, Bi-Liang Lu, Hua-Liang Wei, Lei Chen, Xiao-Hua Li, and Matthias Rätsch

Abstract: Fault diagnosis of rolling bearings is an essential process for improving the reliability and safety of the rotating machinery. It is always a major challenge to ensure fault diagnosis accuracy in particular under severe working conditions. In this paper, a deep adversarial domain adaptation model (called DADA) is proposed for rolling bearing fault diagnosis. This model constructs an adversarial adaptation network to solve the commonly encountered problem in numerous real applications: the source domain and the target domain are inconsistent in their distribution. First, a deep stack auto-encoder (DSAE) is combined with representative feature learning for dimensionality reduction, and such a combination provides an unsupervised learning method to effectively acquire fault features. Meanwhile, domain adaptation and recognition classification are implemented using a Softmax classifier to augment classification accuracy. Second, the effects of the number of hidden layers in the stack auto-encoder network, the number of neurons in each hidden layer, and the hyperparameters of the proposed fault diagnosis algorithm are analyzed. Thirdly, comprehensive analysis is performed on real data to validate the performance of the proposed method; the experimental results demonstrate that the new method outperforms the existing machine learning and deep learning methods, in terms of classification accuracy and generalization ability.

***Index T*erms—fault diagnosis, bearing, feature extraction, stack auto-encoder (SAE), unsupervised learning, domain adaptation, adversarial network, machine learning, deep learning ,deep neural networks.**

I. INTRODUCTION[1]

Rolling bearings are widely used in industrial system, such as wind turbine, aeroengines, and high-speed railways, and it usually plays a pivotal role in their functioning [1]-[4]. However, these devices often work with heavy loads or under some severe environments (e.g., high speed, high humidity, high temperatures and variable speed, etc.), which makes rolling bearings prone to fault attacks. The high failure rate of rolling bearings also increases the operation and maintenance costs. Moreover, in cases in which the potential faults of rolling bearings are not detected, there would be a high risk exists of the breakdown of the entire equipment [5]-[10].Therefore, it is always desirable and necessary to diagnose potential rolling bearing faults in time.

In the Internet of Things (IoT) and Industry 4.0 era, large amounts of real-time data have been collected from the device-monitoring systems. The data, together with modern data mining techniques, makes it possible to effectively mine features and diagnose faults using artificial intelligence methods, such as Support Vector Machine (SVM)[11], Artificial Neural Network (ANN)[12],[13],Stack Auto-Encoder network (SAE)[14], and Deep Belief Network (DBN)[15], [16]. For example, Jiang et al. [13] proposed an approach for rolling bearing fault identification using multilayer deep convolutional neural network. Sun et al.[14] designed an intelligent bearing fault diagnosis method combining compressed data acquisition and deep neural network architecture. Chen et al. [15] presented a novel method to implement bearing fault diagnosis utilizing the integration method of sparse auto-encoder and deep belief network. However, although these intelligent fault diagnosis methods achieve good classification performance in experimental testing, they do not exhibit satisfactory performance when applied in practical applications, in which the classification accuracy is usually much lower than that for test data. This can be explained from two aspects as follows. Firstly, these artificial intelligence methods require a large amount of labeled data to train the model. However, in many real applications, it is very expensive or difficult, even not possible; to collect labeled training data that has the same distribution as the test set. In conclusion, it is difficult to collect sufficient labeled data and then train a reliable diagnosis model in engineering scenarios. Secondly, it is assumed that the training data set and the test set of the model are generated under the same working conditions in the experimental testing. In other words, it is assumed that all data obey the same distribution and possess the same feature space. In reality, however, during the operation of the rotating machinery differ, the mechanical working conditions vary, the signal acquisition methods are different, and the mechanical workloads are varying. As a consequence, these intelligent diagnosis methods have poor generalization ability in the in reality application, and therefore bring poor diagnostic accuracy [11-16].Fortunately, the domain adaptation (DA) technique can be utilized to solve or alleviate the data inconstancy issue (i.e., the inconsistency between the training and test data) [17], [18]. DA aims to reduce the difference between

[1]Manuscript received January 1, 2019; revised May 15, 2019; accepted July 18, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61503134, Grant 61573299, and the Hunan Provincial Young Talents Project under Grant 2018RS3095, Hunan Provincial Natural Science Foundation of China under Grant 13JJ8014.

Z.-H. Liu, B.-L Lu, L. Chen and X.-H Li are with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, 411201, China (e-mail:zhaohualiu2009@hotmail.com, 1197393632@qq.com, chenlei@hnust.edu.cn, lixiaohua_0227@163.com).

H.-L. Wei is with the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield S1 3JD, U.K(w.hualiang@sheffield.ac.uk).

M. Rätsch is with the Image Understanding and Interactive Robotics Group, Reutlingen University, D-72762 eutlingen, Germany (e-mail: matthias.raetsch@reutlingen-university.de)

multi-domains, through learning the invariant knowledge hidden within multiple different domain datasets. So, by using a similar (but not exactly the same) source domain, it provides a solution to the problem of insufficient labeled samples in engineering scenarios. These two aspects make the learned classifiers from the source domain more robust when dealing with mismatched distributions [19]-[21].

DA methods can be divided into two categories: (1) semi-supervised and (2) unsupervised. For example, Pan et al. [22] proposed a transfer component analysis (TCA) technique to reduce the difference of features in two domains. Lu et al. [23] proposed a novel deep model called Deep Model Based Domain Adaptation for Fault Diagnosis (DAFD) in which the AE was combined with domain adaptation. However, these methods are usually achieved by minimizing some predefined distance measures of domain discrepancy such as maximum mean discrepancy (MMD) [24], [25], Bregman divergence [26], or KL-divergence [27]. An advantage of using the existing definition of domain discrepancy such as MMD as a part of the loss function when performing domain adaptation is that the method implementation is simple and straight forward. But such a method has a challenging issue when the data set has a small or limited number of target domain data in the training phase. Thus, these domain adaptation techniques may fail in this case, since they all require using enough target domain data (the number of target domain is similar to the source domain data)to define the domain discrepancies between domains. Inevitably, the lack of target domain data is a frequently encountered scenario in engineering, and an approach of implementing domain adaptation through the predefined distance measures cannot always achieve satisfactory results for such problems. In short, these domain adaptation methods have some detects for fault diagnosis, for example, they can become less effective to define the domain discrepancy distance when the target domain data are not enough.

Recently, deep learning [28] has attracted considerable attention from researchers, since it could capture more hidden knowledge in the process of feature extraction in hierarchical structures. Moreover, integrating the distribution differences in multi-domains, deep learning has well data adaptability in domain adaptation, and possesses strong capabilities in domain-invariant feature learning. In particular, one of the most significant advances in deep learning architecture is the introduction of generative adversarial networks (GANs) [29], which offer strong distribution learning and sample generation ability. The key idea of this method is to train a discriminator and a generator, leading them to Nash equilibrium. In addition, the generator is used to capture the data distribution, and the discriminator is employed to estimate the probability, and the whole training procedure for generator aims to maximize the probability of the discriminator producing the discriminated error. This novel idea provides a way to achieve domain adaptation without extensively target data. On the other hand, the emergence of the stack auto-encoder(SAE) method, which can automatically extract more useful knowledge behind high-dimensional data as a feature learning method with the deep architecture. This characteristic can solve some problems of the shallow structure, such as representative features learning and dimensionality reduction. Finally, combining the novel idea of GANs and SAE can overcome the issues or drawbacks relating to the domain adaptation methods mentioned above.

In this paper, an SAE based deep adversarial domain adaptation (DADA) model is proposed for rolling bearing fault diagnosis. The presented method minimizes an approximate domain discrepancy distance through an adversarial objective with respect to a domain discriminator. Another important advantage of the method is that it employs an unsupervised learning method and can be trained with an end-to-end network. It also offers the following two major advantages, in particular for rolling bearing fault diagnosis:(1) it can effectively solve the data inconsistency issue in which the training data and the test data have inconsistent distribution in the absence of extensive target domain data; and (2)the method can be easily implemented for real applications. To the best of our knowledge, this constitutes the novel work to address the rolling bearing fault diagnosis problem through adversarial networks, and the main contributions are summarized as follows:

1) In order to ensure the reliability of the model when the labeled training data is not enough, we construct a novel deep domain adaptation model based on the GANs, utilizing the sufficient labeled source domain data and then training a reliable diagnosis model in engineering scenarios. A multi-layered network is also used to learn rich knowledge of the source domain to promote domain adaptation.
2) In order to solve the bearing fault diagnosis problem in complex and uncertain environment, we propose a novel deep adversarial domain adaptation algorithm forbearing fault diagnosis based on the model proposed in 1).The adversarial domain adaptation can enhance the generalization ability of the proposed bearing fault diagnosis algorithm, and thus fault diagnosis accuracy can be significantly improved.
3) In order to make the experimental more results robust and generalizable, the proposed algorithm was validated through six domain adaptation situation studies. The effects of the number of hidden layers in the stack auto-encoder network, the number of neurons in each hidden layer, and the hyperparameters of the DADA on the model performance are analyzed. The experimental results demonstrate that our model can produce excellent classification accuracy and possesses strong domain adaptive ability.

The remaining parts of this paper are depicted as follows. The proposed adversarial domain adaptation framework is presented in Section II. In Section III, a novel adversarial domain adaptation algorithm is provided, as well as its optimal solution. Experimental results on different domain adaptation situations are given in Section IV. Finally, a brief summary is presented in Section V.

II.THE PROPOSED ADVERSARIAL DOMAINA DAPTATION FRAMEWORK

Rotating machinery is playing an increasing role in the modern industry. To prevent potential the occurrence of faults and fault propagation, it is critical to monitor the state of the rotating machinery healthy state. Artificial intelligent diagnosis methods, which can efficiently process collected vibration data and automatically obtain diagnosis results, are commonly employed for the condition monitoring and fault diagnosis of rotating machinery health monitoring. It is known that several external factors such as

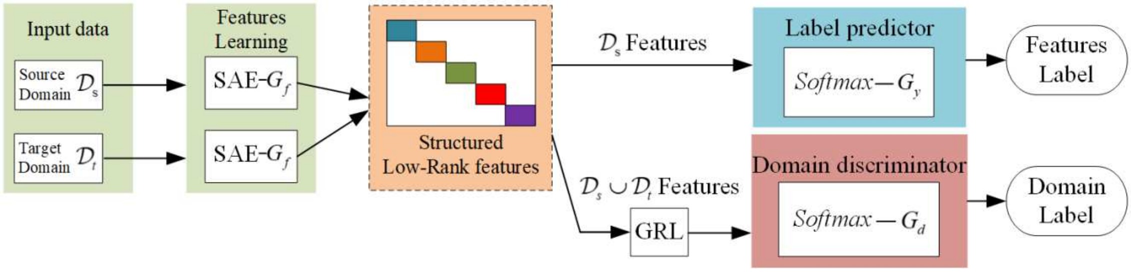


Fig.1. The proposed adversarial domain adaptation framework.

variable working conditions and workloads, can lead to inconsistent distribution in the collected data, which make the traditional artificial intelligence fault diagnosis methods less effective. In recent years, generative adversarial networks (GANs) have attracted increasing interest due to their excellent generative performance. The use of GANs and DA for fault diagnosis can assist to solve the following problem: training data and test data have inconsistent distributions in the rotating machinery. The key idea of GANs is a generator and training of a discriminator, leading them to Nash equilibrium [29]-[31]. GANs make the two networks including the generator model G and the discriminator model D , complementary to each other. Specifically, the former attempts to capture the data distribution, while the latter aims to estimate the probability, and the whole training procedure for G aims to maximize the probability of producing the discriminated error in D .

Similar to GANs, to build a deep adversarial domain adaptation (DADA)framework is as shown in Fig.1, the DADA framework comprises three parts: the generator SAE-G_f to extract features, the domain discriminator G_d implements domain adaptation, and the label predictor G_y obtains feature classification. Note that G_d and G_y can be simultaneously implemented in the proposed adversarial domain adaptation framework, through which fault diagnosis can be achieved by using a similar (but not exactly the same) source domain. The SAE architecture has been extended to comprise several auto-encoder and two Softmax classifiers. This constitutes a general fault diagnosis framework for bearing and a DADA learning process of a two-player game. The first game player is a feature extractor, denoted by SAE-G_f , whose task is to extract the domain-invariant features. The second player is a discriminator-Softmax classifier G_d , which is trained to distinguish whether the features extracted by SAE-G_f belong to the source domain $\mathcal{D}_s$ or the target domain $\mathcal{D}_t$. To extract the domain-invariant feature F , the main purpose of the parameters θ_f in the feature extractor SAE- G_f is to maximize the loss of the domain discriminator G_d and minimize the loss of the label predictor-Softmax classifier G_y .Note that it is impossible to distinguish whether the data comes from the source domain or the target domain only through maximizing the error of the domain discriminator G_d . Domain adaption learning is needed when the two domains become very similar. Features extracted by SAE- G_f (e.g. obtained through minimizing the error of the label predictor) can be used to predict the corresponding label, and to the result in turn can be used for bearing fault diagnosis. By maximizing the error of the domain classifier and minimizing the error of the label predictor in the proposed adversarial domain adaptation model, it is possible to use similar (but not exactly the same) source domain to do fault diagnosis for the target domain data.

The basic function of the adversarial domain adaptation is defined as follows:

$$\begin{aligned} E(\theta_f,\theta_y,\theta_d) = \sum_{X_i\in\mathcal{D}_s} L_y(G_y(G_f(X_i;\theta_f);\theta_y), y_i) - \\ \lambda \sum_{X_i\in\mathcal{D}_s\cup\mathcal{D}_t} L_d(G_d(G_f(X_i;\theta_f);\theta_d), y_i) \end{aligned} \tag{1}$$

where X_i is the training samples; $L_y(\ ,\ )$ is the loss of the label predictor G_y ; $L_d(\ ,\ )$ is the loss of thedomain discriminator G_d ; y_i is the label for X_i ;and λ is a trade-off parameter that controls the proportion of thedomain discriminator loss on the entire loss function. During the training progress, the parameters $\hat{\theta}_f$, $\hat{\theta}_y$,and $\hat{\theta}_d$ deliver a saddle point of the functional (1):

$$(\hat{\theta}_f,\hat{\theta}_y) = \arg\min_{\theta_f,\theta_y} E(\theta_f,\theta_y,\hat{\theta}_d) \tag{2}$$

$$\hat{\theta}_d = \arg\max_{\theta_d} E(\hat{\theta}_f,\hat{\theta}_y,\theta_d) \tag{3}$$

The following rules are used to update the parameters throughout the training process:

$$\theta_f \leftarrow \theta_f - \mu\left(\frac{\partial L_y^i}{\partial\theta_f} - \lambda\frac{\partial L_d^i}{\partial\theta_f}\right) \tag{4}$$

$$\theta_y \leftarrow \theta_y - \mu\frac{\partial L_y^i}{\partial\theta_y} \tag{5}$$

$$\theta_d \leftarrow \theta_d - \mu\frac{\partial L_d^i}{\partial\theta_d} \tag{6}$$

where μ is the learning rate, and these rules can be embedded into the optimization algorithm that using back-propagation such as stochastic gradient descent (SGD)[21]. The factor $-\lambda$ in (4) represents that the training process, which aims to maximize the objection of (3), and this operation is referred to as the gradient reversal layer (GRL). Due to the existence of the GRL, we can rewrite the basic function of the adversarial domain adaptation as follows:

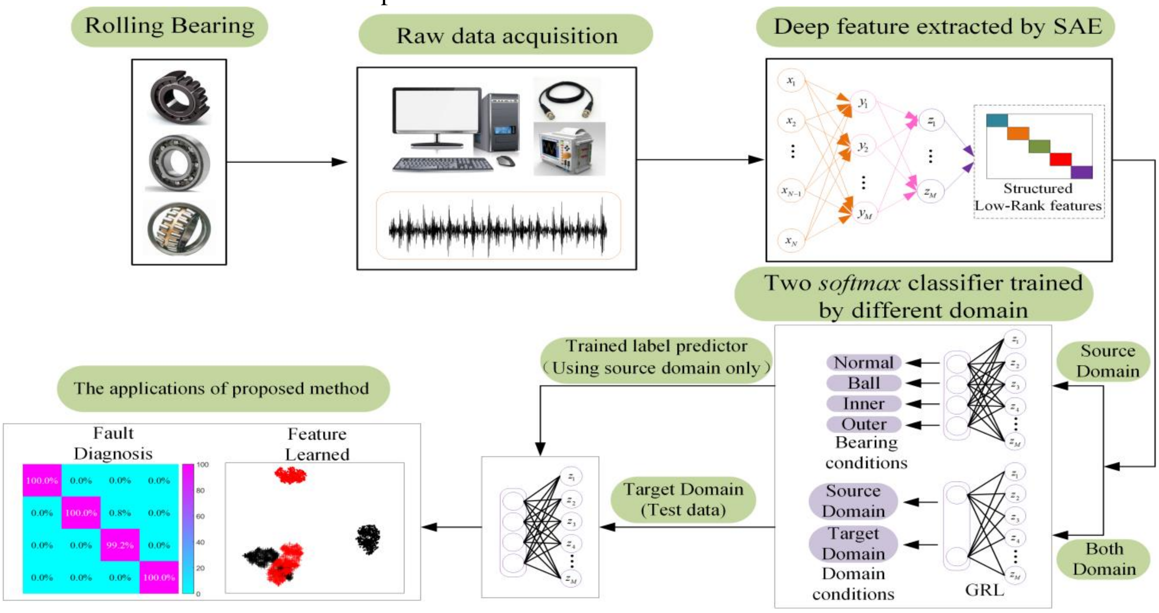


Fig.2. The proposed deep adversarial domain adaptation algorithm for rolling bearing.

$$E(\theta_f,\theta_y,\theta_d)=\sum_{X_i\in\mathcal{D}_s}L_y(G_y(G_f(X_i;\theta_f);\theta_y),y_i)+\lambda\sum_{X_i\in\mathcal{D}_s\cup\mathcal{D}_t}L_d(G_d(G_f(X_i;\theta_f);\theta_d),y_i) \tag{7}$$

Performing SGD or other optimization algorithms on (4)-(6), we can obtain the domain adaptation and sample label prediction simultaneously. Of course, the main difference between DADA and GAN is that the whole training process is distinct. For GAN, the discriminator training and generator training constitute two separate processes. Generally, the generator is trained first, and then the discriminator is trained. For DADA, the training of the discriminator and the training of the generator can be completed in the same training process, and thus the DADA can be trained in an end-to-end framework. The latter is compared with the former, which enables the training process to be implemented in a straightforward manner. In addition, the DADA framework can usually achieve the good domain adaptation ability.

III. THE PROPOSED DEEP ADVERSARIAL DOMAIN ADAPTATION METHOD FOR ROLLING BEARING FAULT DIAGNOSIS

Rolling bearings are a fragile key component of the rotating machinery. Frequently, severe working environments can make rotating machinery vulnerable to damage. Therefore, it is necessary to diagnose potential rolling bearing faults as early as possible. Based on the proposed framework in Section II, in this paper, we propose a deep adversarial domain adaptation algorithm for rolling bearing fault diagnosis. This model consists of three parts: (1) a feature extractor using the stack auto-encoder; (2) the domain discriminator and the label predictor based on the Softmax classifier, and (3) the optimization solution. The approximate algorithm framework is presented in Fig.2, and detailed descriptions of the algorithms are given below.

A. Stack Auto-Encoder -Based Feature Extractor

After the experimental device is operated for a long period of time (e.g.,72 hours) , a large amount of vibration data can be obtained from the rolling bearings [32]-[34], and these raw vibration data contain some noise. Thus, it is required to preprocess the raw vibration data. In this paper, we consider the stack auto-encoder (SAE) [35],[36], which is a useful method to find the representative features of the raw data, because it can reduce the dimension of the collected vibration data and extract high-dimensional features. The main process is depicted in Fig.3.

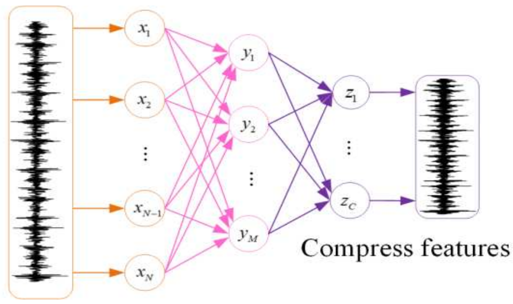


Fig.3. The main process of SAE networks.

The SAE is usually achieved by several auto-encoder (AE) [37] stacked networks stacked. Here, we select two AE networks to stack into the SAE, which is used to extract the representative feature. The reason for such a choice will be explained in the Section IV. The AE network is composed of an encoder and a decoder, which is trained as an unsupervised learning method. Given the input data $X\in\mathbb{R}^N$, the role of the encoder layer is to compress X into the representative feature $Y\in\mathbb{R}^M\,(M<N)$, and the function used as:

$$Y=f(W^{(1)}X+b^{(1)}),\qquad f(x)=\frac{1}{1+\exp(-x)}. \tag{8}$$

where $W^{(1)}$ and $b^{(1)}$ are the weight matrix of size *N*x*M* and bias vector of size M, respectively. $f(x)$ is the activation function of the AE network. Then, the representative feature Y is reconstructed into the vector $\hat{X}$ by the decoder layer as follow:

$$\hat{X} = f(W^{(2)}Y + b^{(2)}) \tag{9}$$

where the $W^{(2)}$ and $b^{(2)}$ are defined in the same way as $W^{(1)}$ and $b^{(1)}$,respectively. The main purpose of AE network training is to obtain $\theta = \{W^{(1)}, W^{(2)}, b^{(1)}, b^{(2)}\}$ by minimizing the reconstruction error between X and $\hat{X}$.

B. The Domain and The Label Discriminator

The Softmax regression model has been widely utilized in fault diagnosis tasks such as fault classification and prediction [38]. In this paper, we use two Softmax classifiers—a label predictor G_y and a domain discriminator G_d .Given the input data $X = \{(x^{(1)}, y^{(1)}), ..., (x^{(m)}, y^{(m)})\}$, with k types of labels $y^{(i)} = \{1, 2, ..., k\}$, let the hypothesis function be $h_\theta(x) = 1/1 + \exp(-\theta^T x)$. The probability that $x^{(i)}$ belongs to each type of label is:

$$h_\theta(x^{(i)}) = \left[p(y^{(i)} = 1 \middle| x^{(i)}; \theta), ..., p(y^{(i)} = k \middle| x^{(i)}; \theta)\right]'$$

$$= \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \dots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \tag{10}$$

The cost function of softmax regression model is:

$$J_\theta = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}\right] + \frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=0}^{n}\theta_{ij}^2 \tag{11}$$

where the symbol $1\{\ .\ \}$ is the indicator function, which is defined as1{a true statement} = 1 and 1{a false statement} = 0; θ is the parameter vector of the Softmax regression model, λ is a trade-off parameter for the weight decay term.

C. Optimization Solution

In this paper, we propose a novel adversarial domain adaptation model for rolling bearing fault diagnosis and choose the SGD algorithm to find the optimal solution. The proposed algorithm is briefly summarized in Algorithm 1. Note that the setting of each parameter in the proposed algorithm is just for demonstration, i.e., a good or better setting of these parameters may be available for real applications. Some details and suggestions are provided in Section IV.

IV. Experimental Test

A. Experimental Data Description

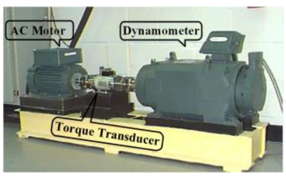


Fig.4. Experimental equipment[39]. Fig.5.The actual situation of three faults.

Algorithm 1

Input: Labeled source data $X_S = \{(x^{(1)}, y^{(1)}), ..., (x^{(n_s)}, y^{(n_s)})\}$ and unlabeled target data $X_t = \{x^{(1)}, x^{(2)}, ..., x^{(n_t)}\}$.

Output: Target class labels.

1.1:Set the SAE parameters:

- learning rate μ = 0.003, weight of sparsity penalty term β =0.0001, sparsity parameter γ =0.1,
- #hidden layer=2,
- #nodes=150 per hidden layer ;

1.2:Set the Softmax classifier parameters:

- Trade-off parameter for the weight decay term λ = 0.0001;

2: Obtain the SAE feature vectors: $\{F_i\}_{i=1}^{n^S+n^t} = trained\_SAE\left(X_S \cup X_t\right)$;

3: Optimize label predictor G_y using SGD on the labeled source features $\{F_i\}_{i=1}^{n^s}$ only, and optimize domain discriminator G_d using SGD on the SAE feature vectors $\{F_i\}_{i=1}^{n^s+n^t}$;the backpropagation of two optimization processes can be written as follow:

$$\theta_f \leftarrow \theta_f - \mu\left(\frac{\partial L_y^i}{\partial \theta_f} - \lambda\frac{\partial L_d^i}{\partial \theta_f}\right),$$

$$\theta_y \leftarrow \theta_y - \mu\frac{\partial L_y^i}{\partial \theta_y},$$

$$\theta_d \leftarrow \theta_d - \mu\frac{\partial L_d^i}{\partial \theta_d}.$$

4: Feed the unlabeled target features $\{F_i\}_{i=1}^{n^t}$ to label predictor G_y and estimate the corresponding labels;

Rolling bearing vibration data were collected from the Case Western Reserve University (CWRU) Bearing Data Center [39], and the equipment is listed in Fig.4. Specifically, the vibration data were obtained by the accelerometers mounted at the end of an induction, containing both normal and fault data, Fig.5 shows the actual situation of the three faults. For each fault, there are four fault sizes corresponding to various fault levels, which are 0.007, 0.014, 0.021 and 0.028. Additionally, the data were acquired at different motor loads (0, 1, 2 and 3 hp) with a sampling frequency of 12 kHz. In this paper, we choose the fault size (0.007) and four different loads(0, 1, 2 and 3 hp) to simulate the scenario for domain adaptation, and it constitutes a total of six domain adaptation scenarios (trial number 1 - trial number 6). The preprocessing procedures were implemented on the row data:6400 samples of 1200 sample length with 80% overlap are selected from both $\mathcal{D}_s$ and $\mathcal{D}_t$.

Finally, the classification accuracy of each method is defined as:

TABLE I
DATA CLASSIFICATION ACCURACY

Methods	0-1hp(Trial number 1)	0-2hp(Trial number 2)	0-3hp(Trial number 3)	1-2hp(Trial number 4)	1-3hp(Trial number 5)	2-3hp(Trial number 6)
Softmax	77.2	66.8	80.6	68.8	60.2	74.0
SVM	93.8	87.7	92.9	74.8	77.9	87.8
BP	94.2	72.1	74.9	65.7	89.2	90.8
SAE	78.6	75.0	90.0	74.9	80.2	75.6
TCA	97.9	85.0	96.8	80.2	94.7	80.5
DAFD	96.7	92.3	97.6	89.4	93.2	92.5
Proposed method	**99.2**	**98.7**	**99.8**	**94.8**	**96.4**	**100.0**

$$accuracy(C\%) = \frac{label(x) = k \cap predict(x) = k}{x_n} \quad (12)$$

where x_n is the total number of test samples, and k is the true label value that a classifier correctly identified.

B. Comparison with the Traditional Methods

Unlike traditional methods (e.g. Softmax, SVM, BP, and SAE), the proposed method focuses on intelligent fault diagnosis with domain shift situation. In addition, the SAE is the case with the proposed method without the adversarial domain adaptation ability. To show the superiority of the proposed method, we compare it with the state-of-the-art domain adaptation methods including TCA [22] and DAFD [23].The application details of these compared methods are summarized as follows: For SVM, Softmax, BP and SAE, the training data of the classification model are derived from the $\mathcal{D}_s$ and $\mathcal{D}_t$, and then select the data of the target domain $\mathcal{D}_t$ are selected as the test data to complete the prediction. The TCA firstly implements domain adaptation and dimensionality reduction for $\mathcal{D}_s$ and $\mathcal{D}_t$ data, and then uses $\mathcal{D}_s$ data to train the SVM to predict the $\mathcal{D}_t$ data labels. The effect of hyper-parameters for each compared method is also empirically analyzed. Due to the space limitations, analysis details are omitted. The accuracies of the seven methods, for a total of six domain adaptation situations, are shown in Table I. For convenience of comparisons, the accuracies are also displayed in histogram format (see Fig. 6). It can be seen that the average accuracy of the proposed method is 98.15%, which is much higher than the six compared methods. It is noticed that the average accuracy of TCA and DAFD is also higher than the other four methods, but the domain adaptation ability of the TCA and DAFD is lower than the proposed method. Fig.7 shows the confuse matrix of the proposed method for the first domain shift situation (trial number 1).

Two points are summarized as follows: (1) the results are obtained through the six domain adaptation scenarios demonstrate that the proposed method can improve the classification ability of the model in a domain shift situation; and (2) the adversarial domain adaptation scheme can produce a superior solution to the domain shift problem compared to traditional domain adaptation methods.

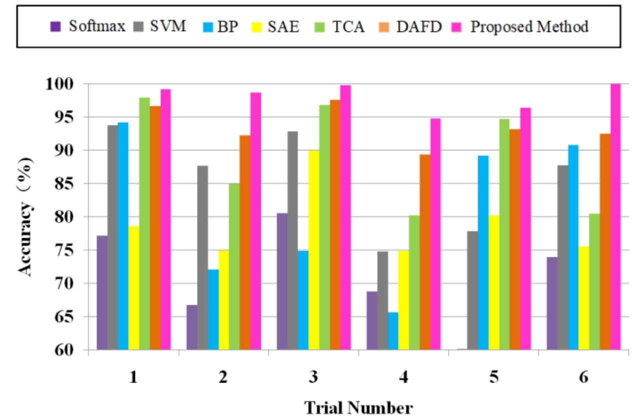


Fig.6. Accuracy on bearing fault diagnosis.

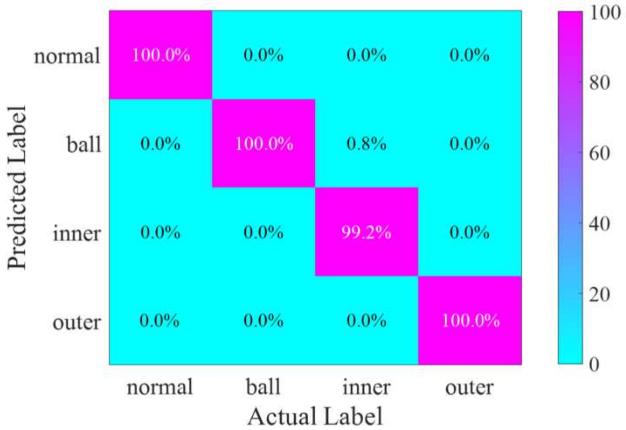


Fig.7. The confusion matrix of the proposed method for the trial number 1.

C. Comparison and Analysis

In order to show the overall performance of the proposed method, TCA is chosen to be the reference method, and comparisons are carried out based on predefined distance such as maximum mean discrepancy (MMD). For TCA, the kernel is defined as Radical Basis Function (RBF), and the optimized subspaces (optimized subspaces are the transformed source and target domain features dimension after adaptation) for the transformed source and target domain features are chosen from $\{8,16,32,64,128\}$.The target domain data for domain adaptation means existing historical data, and test data set are those that need to be diagnosed in real time. As shown in Table II, the target domain data for domain adaptation is gradually reduced(from 700 to 100), but the number of test data set remains unchanged.

TABLE II
DATA CLASSIFICATION ACCURACY

Methods	700	600	500	400	300	200	100
0-1hp (trial number 1) — the number of target domain data for domain adaptation							
TCA	87.63%	83.75%	78.13%	62.58%	62.54%	25.69%	12.59%
Proposed method	98.27%	96.34%	98.50%	97.72%	94.16%	95.08%	86.25%
0-2hp (trial number 1) — the number of target domain data for domain adaptation							
TCA	80.59%	84.85%	75.33%	69.34%	64.08%	18.34%	22.75%
Proposed method	100.00%	92.26%	93.33%	97.50%	96.17%	99.97%	87.09%
0-3hp (trial number 3) — the number of target domain data for domain adaptation							
TCA	88.78%	82.58%	74.52%	66.74%	48.04%	20.13%	25%
Proposed method	96.58%	98.43%	94.38%	90.85%	92.56%	92.48%	75.00%
0-4hp (trial number 4) — the number of target domain data for domain adaptation							
TCA	76.36%	86.35%	72.05%	68.43%	57.47%	18.34%	18.75%
Proposed method	96.84%	84.81%	78.25%	80.94%	85.16%	73.88%	72.22%
0-5hp (trial number 5) — the number of target domain data for domain adaptation							
TCA	89.59%	85.64%	71.60%	64.67%	46.83%	25%	25%
Proposed method	92.25%	98.50%	88.72%	82.75%	75.91%	75.00%	75.78%
0-6hp (trial number 6) — the number of target domain data for domain adaptation							
TCA	78.64%	81.39%	78.14%	56.75%	44.38%	25%	24.84%
Proposed method	98.56%	96.31%	99.81%	99.50%	99.78%	100.00%	74.84%

With the reduction of target domain data for domain adaptation, the performance of TCA drops sharply, especially when the number of target domain data for domain adaptation reaches 100.

It is worthy to stress the following points. Note that TCA can show good performance when the number of source domain is similar to the target domain data, but when additional new data are added and make the source domain and target domain data unbalanced, the performance of TCA drops dramatically. It is known that bearing fault diagnosis needs to process new real-time data in engineering scenarios, as a consequence, methods based on predefined distance such as TCA may fail. The proposed method is superior to the TCA method in that it is more suitable for fault diagnosis of bearings in the engineering scenarios.

D. Empirical Analysis

In order to prove the transferability of the proposed method, the t-distributed stochastic neighbor embedding (t-SNE) software is used to visualize the high-dimensional model and SAE features in a two-dimensional map. All experimental data of this part are taken from the third domain shift situation. t-SNE is a visualization tool that can learn the local structure of high-dimensional data and reduce it to 2-D or 3-D display. Additional details about t-SNE can be found in the [40]. The proposed method can reduce the difference between the source domain and target domain, and Fig. 8(a)-(d) shows this situation.

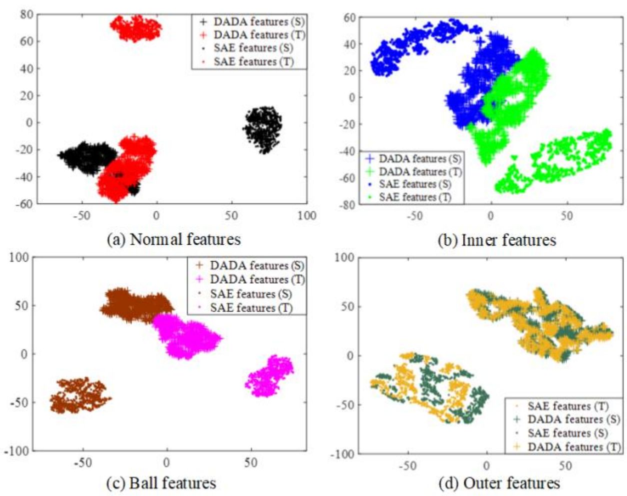


Fig.8. Four results of features displayed by t-SNE dimensionality reduction. The bracketed symbol S represents the source domain $\mathcal{D}_s$,and symbol T represents the target domain $\mathcal{D}_t$. In particular, the SAE features represent the DADA method to remove the domain adaptation ability.

TABLE III

DIAGNOSIS RESULTS BASED ON DIFFERENT HIDDEN LAYER NEURONS

Hidden layer neurons	Average testing accuracy	Time(s)
100 neurons	97.6453%	**36.30**
150 neurons	**98.2930%**	49.40
200 neurons	97.7447%	60.64
250 neurons	96.5606%	76.85
300 neurons	89.7745%	93.80
350 neurons	82.9702%	110.69
400 neurons	80.9427%	130.67

TABLE Ⅳ

DIAGNOSIS RESULTS BASED ON DIFFERENT LEARNING RATES

Learning rate μ	Average testing accuracy
1×10^{-6}	97.5787%
1×10^{-5}	97.6194%
1×10^{-4}	94.8956%
1×10^{-3}	89.8950%
5×10^{-3}	**99.8325%**
1×10^{-2}	99.8277%
1×10^{-1}	99.8113%
5×10^{-1}	25.0000%

Taking the Fig. 8(a) as an example, the distance between the normal features of MODEL (the black cross marker from $\mathcal{D}_s$ and the red cross marker from $\mathcal{D}_t$) is much closer than the SAE features (the black dot marker from $\mathcal{D}_s$ and the red dot marker from $\mathcal{D}_t$).There are similar situations in Fig. 7(b)-(d), which are inner, ball and outer features, respectively. These results explain the excellent domain adaptation capability of the proposed method. Therefore, the classification model trained with the $\mathcal{D}_s$ features can be directly utilized for the classification of $\mathcal{D}_t$ features.

E. Model Structure Analysis

In order to fully explore the potentials of the proposed method, it is essential to analyze the impact of the number of the AE layers and the number of neurons in each layer. In doing this, all input data are taken from the third domain shift situation (trial number 3), and the classifiers are chosen to be Softmax. Fig.9(a) and (b) show the accuracy of SAE, running 200 times in the case of four types of hidden layers, which are 0 hidden layer, 1 hidden layer, 2 hidden layers and 3 hidden layers, respectively. It seems that the accuracy of the 2 hidden layers is higher and more stable. The analysis results for the number of hidden layer neurons are shown in Table III, in which the average classification accuracy was calculated based on the results of 200 iterations. Obviously, although the number of neurons in the hidden layer increases, the average classification accuracy of the proposed method does not increases accordingly. So, we use 150 neurons to form the hidden layer.

F. Model Hyper-parameter Analysis

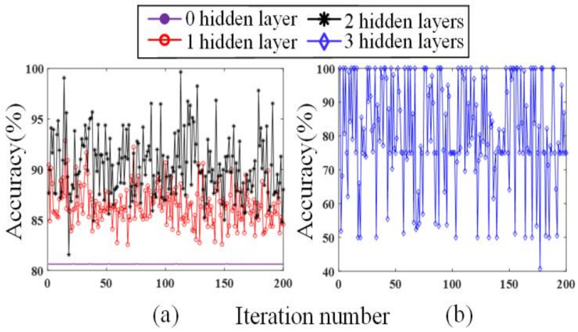


Fig.9. Four kinds of hidden layer analysis results.

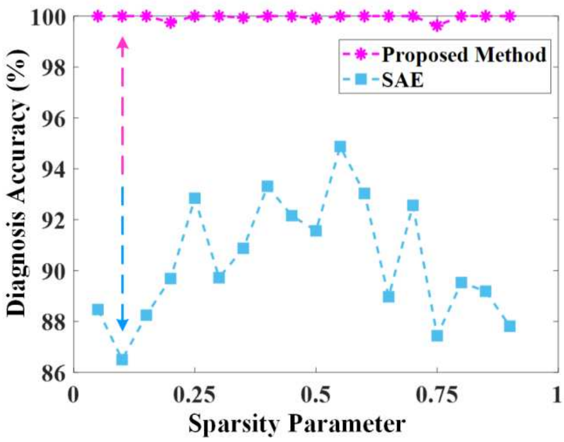


Fig.10. Sparsity parameter analysis results.

In this section, we choose the input data from the third domain shift case (trial number 3) to perform analysis on the sparsity parameter and learning rate, and the classifier is Softmax. Eighteen different values of sparsity parameters are used to detect the classification accuracy of the proposed method and the SAE, the experimental results are shown in Fig.10, where the distance between the blue-dotted line and the red-dotted line represent the accuracy gap between the non-DA and DA methods, respectively. Additionally, experimental results for the learning rate are displayed in Table Ⅳ.

Obviously, $\mu = 5\times10^{-3}$ is the optimal choice. In short, it can be observed that: (1) the proposed method possesses robust classification performance, allowing the parameter to changes in a wide range. The changes of the accuracy still remains in a small band which is much smaller than that of SAE; And (2) the proposed DADA achieves a better classification performance in the domain adaptation situation, demonstrating that it is an effective approach for solving the domain shift problem.

V. Conclusion

This paper proposed a novel model, DADA, to solve the domain shift learning problem in the rolling bearing fault diagnosis field. First, a stack auto-encoder was employed to extract the representation features from the collected vibration data. Second, the label classifier was used to predict the label of the corresponding features. Meanwhile, a domain discriminator was designed and combined with the two sub-network models to construct a whole DADA model.

The proposed method has been applied to real data and its performance has been analyzed. Experimental results demonstrate that the proposed method outperforms the compared peer methods. In addition, the structure and parameters of the model have been analyzed in detail. These analyses are useful for further explore the potential of the associated methods and algorithms.

REFERENCES

[1]W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.

[2]H.David, R.Y Li and J.D Zhu , "Plastic Bearing Fault Diagnosis Based on a Two-Step Data Mining Approach," *IEEE Trans. Ind. Electron*., vol. 60, no. 8, pp. 3429-3440, Aug. 2013.

[3] W. Zhou, B. Lu, T. G. Habetler, and R. G. Harley, "Incipient bearing fault detection via motor stator current noise cancellation using wiener filter,"*IEEE Trans. Ind. Appl*., vol. 45, no. 4, pp. 1309–1317, Aug. 2009.

[4] S. Lu, Q. He, T. Yuan, and F. Kong, "Online Fault Diagnosis of Motor Bearing via Stochastic–Resonance-based Adaptive Filter in an Embedded System", *IEEE Trans. Syst.,Man, Cybern.,Syst.,* Vol.47 , no. 7 , pp.1111 - 1122. Jul. 2017

[5] A. Abid , M. T.Khan, M.S. Khan,"Multidomain Features-Based GA Optimized Artificial Immune System for Bearing Fault Detection",*IEEE Trans. Syst.,Man, Cybern.,Syst.*,,doi: 10.1109/TSMC.2017.2746762,Sep.2017

[6] X. Jin, W. Qiao, Y. Peng, F. Cheng, and L. Qu, "Quantitative evaluationof wind turbine faults under variable operational conditions," *IEEE Trans.Ind. Appl*., vol. 52, no. 3, pp. 2061–2069, Jun. 2016.

[7] B.Y Yang, R.N Liu and X.FChen ,"Fault Diagnosis for a Wind Turbine Generator Bearing via Sparse Representation and Shift-Invariant K-SVD,"*IEEETrans. Ind.Informat*., vol. 13, no. 3, pp. 1321 - 1331, Jun. 2017.

[8] M.Cococcioni, B.Lazzerini and S. L.Volpi,"Robust Diagnosis of Rolling Element Bearings Based on Classification Techniques," *IEEETrans. Ind.Informat*., vol. 9, no. 4, pp. 2256 - 2263, Nov. 2013.

[9] T. Berredjemaand M. Benidirb, "Bearing faults diagnosis using fuzzy expert system relying on an Improved Range Overlaps and Similarity method,"*Expert Systems with Applications*., Vol. 108, pp. 134-142, Oct. 2018.

[10] M. Hamadache,L. Dongik and K. C. Veluvolu, "Rotor Speed-Based Bearing Fault Diagnosis (RSB-BFD) Under Variable Speed and Constant Load," *IEEE Trans. Ind. Electron*., vol. 62, no. 10, pp. 6486 - 6495, Oct. 2015.

[11] Y. Yang, D. Yu, and J. Cheng,"A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM,"*Measurement*.,vol. 40, no. 9–10, pp. 943-950,Nov. 2007.

[12] W. Lu, X. Wang, C. Yang and T. Zhang,"A novel feature extraction method using deep neural network for rolling bearing fault diagnosis,"in*Proc.IEEE Control and Decision Conference*, pp. 2427-2431, May. 2015.

[13] H. Jiang, F. Wang, H. Shao and H. Zhang,"Rolling bearing fault identification using multilayer deep learning convolutional neural network,"*Journal of Vibroengineering*., vol. 19, no.1, pp. 138-149, Feb. 2017.

[14] J.D. Sun, C. Yan and J. Wen, "Intelligent Bearing Fault Diagnosis Method Combining Compressed Data Acquisition and Deep Learning," *IEEE Trans. Instrum. Meas*.,vol. 67, no. 1, pp. 185 – 195, Jan. 2018.

[15] Z.Y ChenandW.H Li, "Multisensor Feature Fusion for Bearing Fault Diagnosis Using Sparse Autoencoder and Deep Belief Network,*IEEE Trans. Instrum. Meas*., Vol. 66, no. 7, pp. 1693 – 1702, July 2017.

[16] H.D Shao, H.K Jiang, H.Z Zhang, T,C Liang, "Electric Locomotive Bearing Fault Diagnosis Using a Novel Convolutional Deep Belief Network,"*IEEE Trans. Ind. Electron*.,Vol. 65,no. 3, pp. 2727 – 2736, March 2018.

[17] S.J. Pan and Q. Yang,"A Survey on Transfer Learning,"*IEEE Trans. Knowl.DataEng*.,vol. 22.no, 10, pp. 1345-1359, Oct. 2010.

[18] K. Weiss, T.M.Khoshgoftaar, D.D. Wang, "A survey of transfer learning," *Journal of Big Data*, Vol. 3, no. 1, pp.9, Dec. 2016.

[19]W.N. Lu,B. Liang, Y. Cheng, D.S.Meng, J. Yang and T. Zhang,"DeepModelbased Domain Adaptation for Fault Diagnosis,"*IEEE Trans. Ind. Electron.*,vol. 64, no. 3, pp.2296 – 2305, March. 2017.

[20] J.YXie, L.B Zhang, L.XDuan, J.J Wang, "On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on Transfer Component Analysis," in *Proc.IEEE Prognostics and Health Management Conference*, Aug. 2016.

[21] L.Wen,L.Gao, and X. Y. 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.136-143,Jan. 2019.*

[22] S.J. Pan,I.W. Tsang, J.T.KwokandQ.Yang,"Domain Adaptation via Transfer Component Analysis,"*IEEE Trans. Neural Netw*.,vol. 22, no.2, pp.199-210, Feb. 2011.

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

[24] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," in Proc. AAAI Conference on Artificial Intelligence, ser. AAAI, 2008.

[25]M.S Long, J.M Wang, J.G Sun, and Philip.S. Yu,"Domain Invariant Transfer Kernel Learning,"*IEEE Trans. Knowl. Data Eng*., vol. 27, no. 6, June 2015.

[26] S. Si, D. Tao, and B. Geng, "Bregman divergence-based regularizationfor transfer subspace learning,"*IEEE Trans. Knowl. Data Eng*., vol. 22, no. 7, 2010.

[27] B. Quanz, J. Huan, and M. Mishra, "Knowledge transfer withlow-quality data: A feature extraction issue,"*IEEE Trans. Knowl. Data Eng*., vol. 24, no. 10, 2012.

[28]Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[29] I. J. Goodfellow, J.Pouget-Abadie, M. Mirza, B. Xu, D.Warde-Farley, S.Ozair, A.Courville and Yoshua Bengio, "Generative Adversarial Networks," in *Proc.NIPS*, pp. 2672-2680, Jun. 2014.

[30] E.Tzeng, J. Hoffman, K.Saenko, T. Darrell, "Adversarial Discriminative Domain Adaptation," in *Proc.IEEE Conference CVPR*, pp.2962-2971, July. 2017.

[31] K.Seeliger, U.Güçlü,L.Ambrogioni,Y.Güçlütürk andM.A.J.vanGerven, "Generative adversarial networks for reconstructing natural images from brain activity,"*NeuroImage*.,Vol. 181, pp. 775-785,Nov. 2018.

[32] Z.W. Gao, C. Cecati and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part II: Fault Diagnosis With Knowledge-Based and Hybrid/Active Approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6,pp. 3768-3774, Jun. 2015.

[33]B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. E. Orchard, andG. Vachtsevanos, "A probabilistic fault detection approach: Application to bearing fault detection," *IEEE Trans. Ind. Electron*., vol. 58, no. 5,pp. 2011–2018, May 2011.

[34] D. He, R. Y. Li, and J. D. Zhu, "Plastic bearing fault diagnosis based ona two-step data mining approach," *IEEE Trans. Ind. Electron*., vol. 60,no. 8, pp. 3429–3440, Aug. 2013.

[35] J.B. Tan,W.N. Lu,J.N. An and X.Q. Wan, "Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder," in *Proc. IEEE Conference 27th CCDC*, pp.4608 - 4613, May. 2015.

[36] R.Thirukovalluru, S. Dixit, R.K. Sevakula, N.K. Verma and A. Salour, "Generating feature sets for fault diagnosis using denoising stacked auto-encoder," in *Proc. IEEE Conference ICPHM*, pp.1 - 7, Aug. 2016.

[37] H.D. Shao,H.K. Jiang,H.W. Zhao andF. Wang, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis,"*Mechanical Systems and Signal Processing*., Vol. 95, pp. 187-204, Oct. 2017.

[38] Y.G. Lei, F.Jia, J. Lin, S.B Xing and S.X. Ding, "An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data,"*IEEE Trans. Ind. Electron*., vol. 63, no. 5, pp. 3137 - 3147, May. 2016.

[39] K. Loparo, Case western reserve university bearing data center.(2013).[Online].Available:http://csegroups.case.edu/bearing datacenter/pages/

[40] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J.Mach. Learn. Res*., vol. 9, no. 2579/2605, p. 85, Nov. 2008.

Zhao-Hua Liu (M'16) He received M.S. degree in computer science and engineering, and the PhD. degree in automatic control and engineering from the Hunan University, China, in 2010 and2012, respectively. He worked as a visiting researcher in the Department of Automatic Control and Systems Engineering at the University of Sheffield, United Kingdom, from 2015 to 2016.

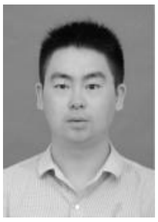
He is currently an Associate Professor of Control and Systems Engineering with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. His current research interests include artificial intelligence and machine learning algorithm design, and condition monitoring and fault diagnosis for electric power equipment. Dr. Liu was a selected as an Excellent Young Talents of Hunan Province of China in 2018. Dr. Liu has published a monograph in the field of *Biological immune system inspired hybrid intelligent algorithm and its applications*, and published more than 30 research papers in refereed journals and conferences, including *IEEE* TRANSACTIONS/JOURNAL/MAGAZINE .He is a regular reviewer for several international journals and conferences.

Bi-Liang Lu He received B. Eng. degree in Electrical engineering and automation from the Hunan university of science and technology, Xiang tan, China, in 2017. He is currently pursuing the M.S. degree in the automatic control and engineering, at Hunan University of Science and Technology, Xiangtan, China.

His current research interests include deep learning algorithm design, and condition monitoring and fault diagnosis for electric power equipment.

Hua-Liang Wei received the Ph.D. degree in the Department of Automatic Control and Systems Engineering, the University of Sheffield, UK, in 2004.

He is currently a Senior Lecturer and Head of the Dynamic Modelling, Data Mining and Decision Making (DM3) laboratory, the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, U.K. His current research interests include system identification and data analytics for complex systems, data driven modeling and data mining, NARMAX methodology and its applications, machine learning and neural networks algorithm design, and parameter estimation.



Lei Chen received M.Sc. degree in computer science and engineering, and the PhD. degree in automatic control and electrical engineering from the Hunan University, China, in 2012 and 2017, respectively.

He is currently a Lecturer with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. His current research interests include deep learning, cloud computing, and big data analysis.

Xiao-Hua Li received the B.Eng. degree in computer science and engineering from the Hunan University of Science and Engineering, Yongzhou, China, in 2007 and the M.Sc. degree in computer science fromHunan University, Changsha, China, in 2010.

She is currently a Lecturer of computer science with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. Her research interest includes evolutionary computation and machine learning.

Matthias Rätsch received Ph.D. degree in the graphics and vision research group(GraVis) at the University of Basel, Switzerland,in 2008.

He is currently a Professor of computer vision and intelligent robots at the image understanding and interactive robotics group, Reutlingen University, Reutlingen, Germany.

His research interests are in the fields of artificial intelligence and machine learning.