# Mechanical fault diagnosis by using dynamic transfer adversarial learning

Yadong Wei[1], Tuzhi Long[1,2], Xiaoman Cai[1], Shaohui Zhang[1,*], Dejan Gjorgjevikj[3], Chuan Li[1]

[1] School of Mechanical Engineering, Dongguan University of Technology, Dongguan 523808, China

[2] School of Automation, Guangdong University of Technology, Guangzhou 510006, China

[3] Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University, Skopje, North Macedonia



**Abstract**: Different machine learning approaches have been developed for the fault diagnosis of mechanical systems. To achieve desired diagnosis performance, lots of labeled one-dimensional signals are required for training machine learning models. However, those signals collected under various working conditions are difficult to be used for both diagnosis model training and testing. For real applications, moreover, the collection of labeled data is more difficult than that of unlabeled ones. To tackle the above challenging points, a dynamic transfer adversarial learning (DTAL) network is proposed for dealing with unsupervised fault diagnosis missions. To this end, an improved feature extractor is developed to deal with one-dimensional mechanical vibration signals. A dynamic adversarial factor is presented to automatically adapt the marginal distribution of the global domain. The conditional distribution of the local domain is employed to make the model independent of training multiple classifiers, so as to reduce the computational burden of the proposed method. The addressed DTAL was evaluated using fault diagnosis experiments for a wind turbine gearbox and benchmark bearings. Compared with other state-of-the-art methods, it has better accuracy and robustness as highlighted by experimental results. The developed model can improve the diagnosis performance under various workloads for mechanical systems.

**Keywords**: Fault diagnosis; dynamic transfer adversarial learning; one-dimensional signal; deep learning; transfer learning.

* Corresponding author. E-mail: zhangsh@dgut.edu.cn

## 1. Introduction

As an essential industry pillar, mechanical systems have been widely used in various scenarios. Due to their long-term operations, machinery aging and failure are inevitable. As the machinery failure may directly lead to severe economic losses and safety accidents, it is vital to diagnose faults accurately.

Some advanced methods have been developed for the mechanical fault diagnosis [1]. Among them machine learning is one of the essential data-driven ways. Such machine learning methods as deconvolution [2], support vector machine [3], and artificial neural network [4] have been proven to be reliable for the fault diagnosis. Considering the fact that there are intense noises in the mechanical signals, deep learning has become a new trend in the study of fault diagnosis towards powerful feature learning and big data processing capabilities. Deep neural networks adaptively capture feature information from original data through multiple nonlinear transformations and approximate complex nonlinear functions with smaller errors [5]. As essential branches of the deep learning, deep enhanced fusion network [6], deep residual network [7], generative adversarial network [8, 9], convolution neural network [10], and recurrent neural network [11] have been widely used in the fault diagnosis. Although those deep learning methods have shown strong diagnostic performance, they require a large amount of labeled one-dimensional signals such as vibration ones for model training. However, it is challenging and time-consuming to collect a large number of labeled samples in practical engineering. Besides, the probability distribution of the collected data changes when the working condition of the mechanical system changes. These factors will inevitably lead to a severe decline in the model's diagnosis performance.

To solve those problems, transfer learning or domain adaptation was developed [12]. The key point of the transfer learning can be used to narrow the data distribution distance between the source domain and the target domain without the target domain data tag through the network training. Zhang et al. [13] presented a domain adaptive model based on convolutional neural network to meet the requirements of fault differentiation and domain invariance. Han et al. [14] designed an intelligent mechanical failure classification framework with deep transport network. A jointly distributed adaptation scheme was used to narrow the difference of data between different domains. Based on the supervised transfer learning of the three-layer sparse autoencoder, Wen et al. [15] proposed a domain adaptive fault diagnosis method by using the maximum average difference term. Guo et al. [16] presented a deep convolutional transfer learning network combining condition recognition and domain adaptation. This method can effectively carry out an unsupervised transfer learning among different mechanical equipment. Li et al. [17] presented a novel intelligent cross-domain fault diagnosis method for rolling bearings. Zhang et al. [18] employed a transfer learning method, which can effectively deal with the big data challenge of faulty samples in data-driven prediction. Xu et al. [19] proposed a two-stage digital twin-assisted fault diagnosis based on deep transfer learning to solve the problem of fault diagnosis in complex industrial manufacturing. These methods have proved that deep learning model can learn more transferable data features and have a combined transfer model with deep learning to extract domain invariant data features for fault diagnosis. Moreover, to reduce the distribution discreteness between different domains, adversarial learning has successfully implemented a migration study function by combining deep learning networks [20]. Guo et al. [21] proposed a generative transfer learning method to solve mechanical fault diagnosis problems under different working conditions. Shao et al. [22] proposed an adaptive method of adversarial domain based on deep transfer learning. To carry out bearing fault diagnosis under different working conditions, Zhang et al. [23] presented a new type of deep transfer model, making use of the property of Wasserstein remote-guided multiple breakthrough network. She et al. [24] proposed a deep multi-feature adversarial transfer diagnosis method based on Wasserstein distance to improved diagnostic performance under different working conditions. Li et al. [25] proposed an adversarial multi-classifier cross-domain fault diagnosis optimization method to improve the accuracy of fault diagnosis by utilizing the overfitting phenomenon of different classifiers in adversarial training. Li et al. [26] designed a deep adversarial transfer learning network by utilizing gradient reversal layer combined with the idea of antagonistic learning. It can detect new faults and realize cross-domain learning. Considering not only the difference of edge distribution between two domains but also the difference of conditional distribution, Jiao et al. [27] designed a novel unsupervised transfer learning framework, referred to as residual joint adaptive adversarial network, for the fault diagnosis. This framework can not only learn classification discrimination for accurate classification but also bridge the edge distribution and joint distribution difference for the domain adaptation. Li et al. [28] proposed an adaptive method in mechanical fault diagnosis based on deep learning. This method can solve the problem of fault migration between different devices.

Those advanced adaptive methods have achieved outstanding diagnostic performance in the field of the fault diagnosis. They either use a single domain discriminator to align the edge distribution (global distribution) between the source domain and the target domain or use multiple discriminators to match the conditional distribution (local distribution). Besides, they may also consider edge distribution and conditional distribution equally important without considering their relative importance. However, the contribution of marginal distribution (global distribution) and conditional distribution (local distribution) is often different in the transfer learning [29]. As shown in Fig. 1, for example, migrating from the source domain to the target domain I indicates that the two domains are incredibly critical. This means that global distribution is critical. Migrating from the source domain to the target domain II indicates that the global distribution is very close. This means that the local distribution should contribute more to the adaption. Besides, migrating from the source domain to the target domain III shows that the two kinds of probability distribution state of chaos. It does not know the probability distribution is more critical.

/*** Insert Figure 1 here ***/

For this reason, this paper proposes a new unsupervised dynamic transfer adversarial learning (DTAL) network to solve those problem and realize more effective transfer diagnosis. On the one hand, to our knowledge, this is the first attempt to dynamically adjust the weight between conditional distribution and edge distribution in the fault diagnosis community. On the other hand, an improved feature extractor is proposed to learn characteristics from one-dimensional vibration signals collected from mechanical systems. Through comprehensive analysis of network visualization and comparison of experimental results, it shows the superiority and strong stability of this method.

The remaining of this paper is structured as follows. In Section 2, the definition of the problem and various structural components of DTAL is detailed. Experiments are introduced in Section 3. Experimental results and comparisons to peer methods are given in Section 4. Conclusions are drawn in Section 5.

## 2. Methodology

The purpose of the transfer diagnosis for the mechanical systems is to classify unlabeled target conditions by transferring the trained source domain data. Suppose that the source domain $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$ has $n_s$ labeled samples and the target domain $D_t = \{x_j^t\}_{j=1}^{n_t}$ has $n_t$ unlabeled samples. Let that label types for the source domain are the same as those of the target domain, i.e., $x_i, x_j \in R^d$ where $d$ is the dimensionality. As the marginal distributions of the source domain and the target domain are different, i.e., $P_s(x_s) \neq P_t(x_t)$, the goal of the proposed dynamic transfer adversarial learning (DTAL) is to design a deep neural network with the function of transfer classifier $y = f(x)$ to reduce the distribution distance between the two domains. This deep neural network can achieve better performance in the target domain by training the labeled source domain and the unlabeled target domain. The details of the addressed DTAL for the mechanical fault diagnosis are given in the following subsections.

### 2.1. Adversarial principle

Following the way of generative adversarial networks (GANs) [22], this work defines $G_d$ as domain discriminator, $G_f$ as feature extractor, and $G_y$ as label classifier. The purpose of training $G_d$ is to identify the source domain from the target one. On the other hand, the purpose of training $G_f$ is to recognize and to extract the invariant features in the domain. This behavior is used to confuse $G_d$. In the training process, let $\theta_f$, $\theta_y$ and $\theta_d$ represent the parameters of $G_f$, $G_y$, and $G_d$, respectively. Since the training processes for $G_d$ and $G_f$ are adversarial, $\theta_f$ is obtained by maximizing the loss of $G_d$ while $\theta_d$ is trained by minimizing the loss of $G_d$. During this process, the loss of $G_y$ is also minimized. Therefore, the loss function $L(\theta_f,\ \theta_y,\ \theta_d)$ can be defined as

$$L(\theta_f, \theta_y, \theta_d) = \frac{1}{n_s} \sum_{x_i \in D_s} L_y\left(G_y\left(G_f(x_i)\right), y_i\right) - \frac{\lambda}{n_s + n_t} \sum_{x_i \in (D_s \cup D_t)} L_d(G_d(G_f(x_i)), d_i) \quad (1)$$

where $\lambda$ is the parameter to adjust the loss ratio of $G_d$, $d_i$ represents a logic flag corresponding to the source domain (0) or the target domain (1), $L_y$ denotes the loss function of the feature extractor, and $L_d$ stands for the loss function of the domain discriminator. Three parameters $\hat{\theta}_f$, $\hat{\theta}_y$, and $\hat{\theta}_d$ form a solution through the adversarial training, given by

$$\left(\hat{\theta}_f, \hat{\theta}_y\right) = argmin_{\theta_f, \theta_y} L(\theta_f, \theta_y, \theta_d) \tag{2}$$

$$\left(\hat{\theta}_d\right) = argmax_{\theta_d} L(\theta_f, \theta_y, \theta_d) \tag{3}$$

## 2.2. Network components of DTAL

Existing research has shown that aligning both the marginal and conditional distributions could produce better result [30]. However, in reality, it is very difficult to explain the relationship between the marginal distribution and the conditional one due to their different effects on domain differences. Therefore, dynamic transfer learning algorithm is presented to solve this problem. This is inspired by manifold embedded distribution aligndment (MEDA) [31] and dynamic adversarial adaptation network (DAAN) [32]. MEDA was proposed to solve the problem of edge and conditional distributions by training linear classifiers in the process of each iteration. Hence, MEDA algorithm needs a lot of time to complete the training and cannot be applied to large datasets. On the other hand, DAAN is capable of solving the large dataset problem.

As shown in Fig. 2, the proposed DTAL is based on the DAAN combined with the idea of MEDA. The basic structure of the DTAL includes a feature extractor, a label classifier, a global domain discriminator, multiple local domain discriminators, and a dynamic adversarial factor. All the DTAL components are detailed in the following subsections.

/*** Insert Figure 2 here ***/

**(1) Feature extractor**

As an important component for dynamic transfer learning, the feature extractor $G_f$ has a big impact on the accuracy of the final classification result. Various classification models such as VGGNet [33], GoogleNet [34], and ResNet [35] has been reported in the in the field of image processing. The feature extractor in this work is inherited from [12] and is improved in this work by extracting and analyzing the feature information of each layer in it.

As shown in Fig. 3, the improved feature extractor is composed of six convolution layers and one fully connected layer. Each convolutional layer carries a MaxPool. The activation function for the feature extractors is ReLU. The improvements of our feature extractor are mainly in two aspects: (1) increase the convolution kernel size of the first convolutional layer; and (2) remove AdaptiveMaxPool and increase the number of MaxPool, to accommodate one-dimensional signals for the mechanical fault diagnosis.

/*** Insert Figure 3 here ***/

The effect of convolutional neural network (CNN) on one-dimensional vibration signal in the field of fault diagnosis differs from that of two-dimensional image data in the field of image [36]. For a 224×224 image in ImageNet [37], although the two-dimensional image performs well under the action of the 3×3 convolution kernel of VGGNet [33], for a 2048×1 one-dimensional vibration signal, it is unrealistic to use 3 × 1 convolution kernel. If only 3×1 convolution kernel is used in the diagnosis model, the network structure will be too deep, the calculation cost will be increased, and the feature extraction process will be easily interfered by the common high-frequency noise in the industrial environment. Therefore, to obtain useful information in the low frequency range of the mechanical vibration signal, a 64×1 wide convolution kernel is first used to extract features in the first convolution layer, and then a continuous 3×1 small convolution kernel is presented to obtain better feature representation. On the other hand, the standard feature extractor given by [12] has defects in using the AdaptiveMaxPool layer in the last convolutional layer. Assuming the shape of the input data is [50, 1, 1024], the shape before the AdaptiveMaxPool of the original model is [50, 128, 500], and the shape after AdaptiveMaxPool is [50, 128, 4]. Excessive pooling power leads to too much loss of information. This is not conducive to feature extraction. The AdaptiveMaxPool used in [12] is only suitable for datasets with small data shapes. Therefore, on the basis of the original model, we use the MaxPool layer with parameter kernel_size = 2, stride = 2 (the parameter name is the same as the name in PyTorch) for each layer of convolutional layer to pool it, ensuring that in the feature extraction process, valid feature information will not be lost in large amounts.

**(2) Label classifier**

The function of the label classifier $G_y$ proposed as shown in Fig. 2 is to classify labels after training the labeled samples in the source domain. In this work, the loss function used by the label classifier is a cross-entropy loss function described as

$$L_y = -\frac{1}{n_s}\sum_{x_i \in D_s}\sum_{c=1}^{C} P_{x_i \to c} log G_y(G_f(x_i)) \tag{4}$$

where $P_{x_i \to c}$ is the probability that $x_i$ belongs to the $c$-th category.

**(3) Global domain discriminator**

As shown in Fig. 2, the function of the global domain discriminator $G_d$ is to align the global distribution of both the source domain and the target domain. The loss function of $G_d$ is defined as

$$L_g = \frac{1}{n_s + n_t}\sum_{x_i \in D_s \cup D_t} L_d(G_d(G_f(x_i)), d_i) \tag{5}$$

where $d_i$ is the domain label of the input sample $x_i$.

**(4) Local domain discriminator**

Local domain discriminator $G_d^c$ is proposed to align the conditional distribution of the source domain and the target domain. Because it is more aligned than the global domain discriminator, there are many patterns of distributions. It can adapt the domain in more fine-grained way. In a nutshell, $C$ categories of the domain discriminator $G_d^c$ can constitute a global domain discriminator. Each local domain discriminator is responsible for the domain classification of its corresponding class. The loss function of the local domain discriminator is formulated as

$$L_l = \frac{1}{n_s + n_t}\sum_{c=1}^{C}\sum_{x_i \in D_s \cup D_t} L_d^c(G_d^c(\hat{y}_i^c G_f(x_i)), d_i) \tag{6}$$

where $L_d^c$ is a cross entropy loss function associated with class $c$, $\hat{y}_i^c$ is the predicted probability distribution over the class $c$ of the input sample $x_i$, and $d_i$ is the prediction domain label or domain label of the input sample $x_i$.

**(5) Dynamic adversarial factor**

In this work, a simple yet effective way is proposed to obtain dynamic adversarial factor $\omega$. At first, the global domain distribution is regarded as the marginal distribution, and the local domain distribution is regarded as the conditional distribution. The weights of the two distributions are obtained using proxy A-distance [38]. In this way, the global A-distance $d_{A,g}(D_s, D_t)$ and local A-distance $d_{A,l}(D_s^c, D_t^c)$ are defined as

$$d_{A,g}(D_s, D_t) = 2(1 - 2(L_g)) \tag{7}$$

$$d_{A,l}(D_s^c, D_t^c) = 2(1 - 2(L_l^c)) \tag{8}$$

By calculating the global distance and local distance, one can finally obtain $\omega$ as

$$\hat{\omega} = \frac{d_{A,g}(D_s, D_t)}{d_{A,g}(D_s, D_t) + \frac{1}{c}\sum_{c=1}^{C} d_{A,l}(D_s^c, D_t^c)} \tag{9}$$

Compared with MEDA, the above equation does not need to train multiple classifiers. The computation burden for obtaining $\omega$ in this work is much less than MEDA.

The initial value of $\omega$ is set as 1 in the first epoch. After each epoch, the label classifier will assign a prediction label to each sample in the target domain. The local distance of each $c$ class is therefore calculated by

$$L_l^c = CrossEntropy(\hat{d}^c, d^c) \tag{10}$$

where $\hat{d}^c = [\hat{d}_s^c; \hat{d}_t^c]$ is the predicted value of the output of the class $C$ domain discriminator. $d^c = [0; 1]$ is the real domain label, where $0 \in R^{|\hat{d}_s^c| \times 1}$ and $1 \in R^{|\hat{d}_t^c| \times 1}$, i.e., the source domain label is 0 and the target domain label is 1.

## 2.3. Development of DTAL for the mechanical fault diagnosis

With all the aforementioned network components, the overall loss functions can be given by combining different component loss functions as

$$L(\theta_f, \theta_y, \theta_d|_{c=1}^{C}) = L_y - \lambda((1-\omega)L_g + \omega L_l) \tag{11}$$

where the proportion of the common influence of $L_g$ and $L_l$ on the overall loss function can be adjusted by $\lambda$. Although $\omega$ is also a hyperparameter, it can be calculated by the neural network itself. When $\omega$ approaches 0, however, the global distribution alignment is the domination (Target I as shown in Fig. 1). When $\omega$ approaches 1, on the other hand, the local subdomain distribution of categories is the domination (Target II as shown in

Fig. 1).

For real applications, the marginal and conditional distributions are not pre-defined. Through dynamic adversarial factor $\omega$, our DTAL network can adapt to different working conditions for dealing with the fault diagnosis tasks. In the network training, parameters to be transferred are defined as $\mathbf{\Theta} = \{\theta_f, \theta_y, \theta_d, \theta_d^c|_{c=1}^C\}$. The gradient of the overall loss function is therefore given by

$$\Delta_{\mathbf{\Theta}} = \frac{\Delta L_y}{\Delta \mathbf{\Theta}} - \lambda \frac{\Delta((1-\omega)L_g + \omega L_l)}{\Delta \mathbf{\Theta}} \tag{12}$$

With this transfer learning procedure, overall steps of the present DTAL for the machinery fault diagnosis are given in Fig. 3 and are detailed as below.

**Step 1**: Collect raw signals from sensors installed on the mechanical system.

**Step 2**: Divide the collected data into source domain data and target domain ones under different working conditions of the machinery.

**Step 3**: Construct the feature extractor, label classifier, global domain discriminator and local domain discriminator using equations illustrated in the subsection 2.2.

**Step 4**: Use the source domain datasets for pre-training the model.

**Step 5**: Unlabeled target domain datasets and labeled source domain dataset are employed for the model transferring. The trained model is employed for the fault diagnosis. End.

/*** Insert Figure 4 here ***/

## 3. Experiments

Two fault diagnosis experiments were carried out for evaluating the present DTAL model. This first case was from our wind turbine diagnosis experiment, and the second one was from a benchmark experimental dataset. As different experiments have different degrees of difficulty in terms of the mechanical fault diagnosis, these two experiments can comprehensively explore the effectiveness and stability of the proposed method.

### 3.1. Experimental configurations

The first experiment was carry out to diagnose the health condition of a wind turbine.

As shown in Fig. 5, a wind turbine (RCVA-3000) was driven by a wind blower to generate electricity through the transmission of a gearbox. Three accelerometers were installed on the gearbox housing to collect vibration signals at the sampling frequency of 100 kHz.

/*** Insert Figure 5 here ***/

The gearbox of the wind turbine consisting of a sun gear, a ring gear, and three planetary gears are also shown in Fig. 5. Different faults were pre-planted to the gearbox as shown in Table 1. A total of six conditions with one healthy and 5 different fault patterns were employed in the experiment. By changing the working load of the wind turbine (and hence the gearbox), the dataset consists of three different operating conditions (workloads 1, 2 and 3). In this way, six transfer learning setups can be obtained through permutation and combination for the dataset collected in this experiment.

/*** Insert Table 1 here ***/

In addition to the fault diagnosis for a wind turbine, a benchmark dataset was employed in the second experiment to validate the effectiveness of the present method. This benchmark dataset was from the public bearing fault diagnosis dataset of Paderborn University [39], collected from a bearing test rig consisting of an electric motor, a torque measurement shaft, a rolling bearing, a flywheel and a load motor. Vibration signals were collected by an accelerometer attached to the bearing housing with the sampling frequency 64 kHz. Bearing faults are classified into 13 categories according to the location, degree, combination, and characteristics of bearing damage. Four different bearing operating conditions, generated by various combinations of two different rotary speeds, two different load torques and two different radial forces, were applied in this experiment. To study transfer learning tasks, there were 13 classes of bearing faults under 4 different operating conditions to be diagnosed. For transfer tasks, task 0 to task 1 (denoted by "0-1") means that the data of source domain were collected under the working conditions with rotation speed 1500 rpm, load torque 0.7 Nm and radial force 1000 N, and the data of the source domain and the target domain were collected under the working conditions with rotary speed 900 rpm, load torque 0.7 Nm and radial force 1000 N. In total, 12 transfer tasks can be set for the benchmark dataset.

### 3.2. Data processing and parametric settings

In the experiments, original vibration signals of the two datasets, i.e., wind turbine dataset and benchmark dataset, were collected at sample frequencies of 100kHz and 64kHz, respectively. To reduce the computational burden, the wind turbine dataset was down-sampled from 100kHz to 25kHz. Original signals of all datasets were divided into 1024 samples without any overlapping and were sent directly to the model input layer as time-domain data. There were 2880 samples collected from the wind turbine gearbox experiment (6 classes each of which had 480 samples), and 6240 samples from the benchmark bearing experiment (13 classes each of which had 480 samples). A z-score normalization was used to keep the input dataset $\mathbf{x} = x_1, x_2, \ldots, x_N$ in a certain range, depicted as

$$x_i^{normalize} = \frac{x_i - x_i^{mean}}{x_i^{std}}, i = 1,2, \ldots, N \tag{13}$$

where $x_i^{mean}$ and $x_i^{std}$ are the mean value and the standard deviation of $x$, respectively.

To avoid the test leakage, 80% of total samples without any overlapping were regarded as the training set, and the remaining samples as the test set in the source and the target domains, respectively. This is illustrated in Fig. 6.

/*** Insert Figure 6 here ***/

For each experiment, the training set was fed to DTAL. Only samples from the source domain were used to train the model. The trained model was directly applied to test the samples of target domains. That means that no samples of target domains participated in the model training, while the source and target domains shared the same model structure and parameter settings. Network structure parameters of the improved feature extractor in the developed DTAL used in the experiments are listed in Table 2. The label classifier is specified in Table 3, and the network structure parameters of the global domain discriminator and local domain discriminator are given in Table 4.

/*** Insert Table 2 here ***/

/*** Insert Table 3 here ***/

/*** Insert Table 4 here ***/

### 3.3. Comparison methods

To demonstrate a comprehensive performance evaluation for the proposed DTAL method, several state-of-the-art deep unsupervised domain adaptation methods, including adaptive batch normalization (AdaBN) [40], multi kernels maximum mean discrepancy (MK-MMD) [41], joint maximum mean discrepancy (JMMD) [30], domain adversarial neural network (DANN) [42], conditional domain adversarial network (CDAN) [43] and source domain convolutional neural network (SCNN). To highlight the performance of the improved feature extractor, an original feature extractor as introduced in Ref. [12] was used to replace the improved feature extractor of the DTAL. This was named as DTAL with standard feature extractor (DTAL-SFE).

For the comparison methods, each of them is composed of a feature extractor and its corresponding migration method, with its structure and parameter settings same as those in its related literature. Besides, a basic method (Basis) was constructed by combining feature extractor and the label classifier to make an evaluation benchmark. Then, only samples from the source domain were used to train the model. Please note that the improved feature extractor is not used in comparison methods All the training and testing data for the comparison methods were the same as the experimental data for DTAL.

All the above mentioned fault diagnosis methods were realized on PyTorch framework [44]. To make the experimental results more reasonable, the feature extractor structures of all methods were set as the same. During the training process, the mini-batch stochastic gradient descent (SGD) with momentum of 0.9 and batch size of 32 were taken as an optimization scheme for the backpropagation. Each experiment related to all methods was trained for 200 epochs. In the first 50 epochs they were pre-trained only with source samples. The initial learning rate was set as 0.01 with a decay (multiplied by 0.1) in the epoch 150. All experiments were carried out under Ubuntu 16.04 and PyTorch 1.3 running on a computer with an Intel Xeon E5-2620 v4, TITAN Xp, and 12GRAM. Classification accuracy on test dataset shown in Eq. (1) was used as the evaluation metric, which was widely used in [45-47] The label of the target domain was only used in the test stage. To reduce the randomness, each experimental result was the average classification accuracy of five experiments for each method.

$$Accuracy = \frac{|x : x \in D_t \wedge \hat{y}(x) = y(x)|}{|x : x \in D_t|} \tag{14}$$

## 4. Results and discussion

### 4.1. Fault diagnosis results

The fault diagnosis results of the proposed method and the comparison ones for the wind turbine gearbox experiment are listed in Table 5. For all the transfer tasks, the present DTAL has the best performance. Moreover, DTAL-SFE model that uses the original feature extractor is compared with other methods. For the first three transfer tasks, DTAL-SFE are better than other methods. Although the latter three transfer tasks donot get the first place, they are all acceptable. The first JMMD model has higher accuracy in the last three transfer tasks, but has a large deviation compared with the accuracy of the first three transfer tasks. The average accuracy of DTAL-SFE is also higher than that of models other than the DTAL model. This shows that the present dynamic transfer adversarial network (even if without the improved feature extractor) can obtain better fault diagnosis performance and strong robustness. In addition, the improved feature extractor can further improve the fault diagnosis accuracy.

/*** Insert Table 5 here ***/

The fault diagnosis results of the proposed method and the comparison ones for the benchmark bearing experiment are listed in Table 6. Once again, the best performance was achieve by the proposed DTAL method. Compared to DTAL-SFE, DTAL has the least accuracy increase of 6.32% (transfer task 0-2), the maximum accuracy increase 39.74%

(transfer task 1-2), and the average accuracy increases 25.15%. This also further indicates the improved feature extractor can improve diagnostic performance.

/*** Insert Table 6 here ***/

As can be seen from Tables 5 and 6, the best two models in the experiments are DTAL and DTAL-SFE, followed by JMMD and CDAN. The good diagnostic performance of JMMD and CDAN is mainly due to the fact that they all consider the edge distribution and conditional distribution between the source and target domains with the same weight. This demonstrates that both edge distribution and conditional distribution can have a significant impact on diagnostic performance in the transfer learning. Our method has the best diagnosis accuracy mainly because of the dynamic factors dynamically adjusting the weight between the edge distribution and the condition distribution. This can be more suitable for mechanical fault diagnosis under different working conditions in practical engineering. The results of two experiments demonstrate that the proposed method has strong diagnosability and generality for the mechanical fault diagnosis.

### 4.2. Discussion on the improved feature extractor

The feature extractor is an important part of the intelligent fault diagnosis model. Its function is to extract features of the dataset for the lower network. It plays a great role in the accuracy of the final model output. To further demonstrate the advantages, the proposed improved feature extractor was compared with the original feature extractor [12] in two aspects: (1) The proposed method and the compared methods both use the improved feature extractor to carry out the diagnosis experiment on the gearbox data set; and (2) In the experiment of bearing data set diagnosis, the feature maps of each layer of the original feature extractor and the improved feature extractor are visualized and compared. In this subsection, the improved feature extractor was used to replace the original one. The compassion methods are therefore named as SCNN-IFE, AdaBN-IFE, MK-MMD-IFE, JMMD-IFE, DANN-IFE, and CDAN-IFE, respectively.

Table 7 shows the fault diagnosis results of the present DTAL and the comparison methods for the wind turbine gearbox. As illustrated by the mean value of the diagnostic accuracies, all methods have good results, and the best one is still DTAL. By comparing with the diagnosis results in Table 5, it is found that the improved feature extractor has a great improvement compared to the initial feature extractor. This proves that the present improved feature extractor can be applied to all methods, and has a certain degree of stability and versatility.

/*** Insert Table 7 here ***/

In addition to the quantitative evaluation as shown in the above table, a visualized comparison was performed using the benchmark bearing dataset. Under the premise of the bearing dataset and the transfer learning task (source domain 0 to transfer to target domain 3), the output between the original feature extractor and each layer of the feature extractor are shown in two-dimensional clustering diagrams as show in Fig. 7, where '×' represents the data of the target domain, and '·' represents the data of the source domain.

Figs. 7 (a)-(h) indicate the clustering results from the MaxPool1d layer of Layer 1, the MaxPool1d layer of the Layer 2, the MaxPool1d layer of Layer 3, the MaxPool1d layer of Layer 4, Layer 5 of the improved feature extraction layer, the MaxPool1d layer of Layer 6, the Linear layer of Layer 7, and the Linear layer of the source_fc layer, respectively. On the contrary, the clustering results for the initial feature extractor are shown in Fig. 7 (i)-(o), respectively. Details for the initial feature extractor are available in Ref. [12]. Comparing the two extractors indicates that the improved feature extractor has better clustering performance.

/*** Insert Figure 7 here ***/

## 5. Conclusions

In this paper, a dynamic transfer adversarial learning (DTAL) has been reported to dealing with mechanical fault diagnosis. Dynamic network structure and improved feature extraction were combined to overcome the transfer problem of different data distributions between the source domain and the target domain, so as to obtain better diagnostic performance for mechanical fault diagnosis under different working conditions. Two typical mechanical systems, i.e., a wind turbine gearbox and a bearing benchmark, were employed to validate the effectiveness of the present DTAL method. The improved feature extraction approach was compared to the standard feature extractor using different experimental data. Thanks to the improved feature extractor, the diagnosis accuracy of all the methods has been significantly improved, which shows that the improved feature extractor is effective and stable for one-dimensional vibration signal processing in the mechanical fault diagnosis. The suggested DTAL was compared to some state-of-the-art peer methods. On the basis of systematic comparative study, DTAL shows the best performance for all the experiments. There are two main contributions in this paper. On the one hand, a new type of unsupervised dynamic adaptive adversarial network for intelligent fault diagnosis was proposed. On the other hand, an improved feature extraction approach was developed to dealing with one-dimensional vibration signals collected from the mechanical systems for the fault diagnosis.

## Acknowledgements

This research is partially supported by the National Natural Science Foundation of China (51975121, 51775112), the Guangdong Basic and Applied Basic Research Foundation (2019B1515120095), and the MoST International Cooperation Program (6-14).

## References

[1] Dai X, Gao Z. From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. IEEE Transactions on Industrial Informatics. 2013, 9, 2226-38.

[2] Miao Y, Zhao M, Lin J, Lei Y. Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings. Mechanical Systems and Signal Processing. 2017, 92, 173-95.

[3] Li C, Cabrera D, Sancho F, Cerrada M, Sánchez R-V, Estupinan E. From fault detection to one-class severity discrimination of 3D printers with one-class support vector machine. ISA transactions. 2020.

[4] Samanta B. Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. Mechanical Systems and Signal Processing. 2004, 18, 625-44.

[5] Jia F, Lei Y, Lin J, Zhou X, Lu N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mechanical Systems and Signal Processing. 2016, 72, 303-15.

[6] Pu Z, Li C, Zhang S, Bai Y. Fault Diagnosis for Wind Turbine Gearboxes by Using Deep Enhanced Fusion Network. IEEE Transactions on Instrumentation Measurement. 2020, 70, 1-11.

[7] Zhang W, Li X, Ding Q. Deep residual learning-based fault diagnosis method for rotating machinery. ISA transactions. 2019, 95, 295-305.

[8] Li C, Cabrera D, Sancho F, Sánchez R-V, Cerrada M, Long J, et al. Fusing convolutional generative adversarial encoders for 3D printer fault detection with only normal condition signals. Mechanical Systems and Signal Processing. 2020, 147, 107108.

[9] Li C, Cabrera D, Sancho F, Sanchez R-V, Cerrada M, de Oliveira JV. One-shot fault diagnosis of 3D printers through improved feature space learning. IEEE Transactions on Industrial Electronics. 2020.

[10] Wen L, Li X, Gao L, Zhang Y. A new convolutional neural network-based data-driven fault diagnosis method. IEEE Transactions on Industrial Electronics. 2017, 65, 5990-8.

[11] Liu H, Zhou J, Zheng Y, Jiang W, Zhang Y. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. ISA transactions. 2018, 77, 167-78.

[12] Zhao Z, Zhang Q, Yu X, Sun C, Wang S, Yan R, et al. Unsupervised Deep Transfer Learning for Intelligent Fault Diagnosis: An Open Source and Comparative Study. arXiv preprint arXiv. 2019, 1912, 12528.

[13] Zhang B, Li W, Li X-L, Ng S-K. Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks. IEEE Access. 2018,

6, 66367-84.

[14] Han T, Liu C, Yang W, Jiang D. Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application. ISA transactions. 2020, 97, 269-81.

[15] Wen L, Gao L, Li X. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. IEEE Transactions on Systems, Man,Cybernetics: Systems. 2017, 49, 136-44.

[16] Guo L, Lei Y, Xing S, Yan T, Li N. Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. IEEE Transactions on Industrial Electronics. 2018, 66, 7316-25.

[17] Li X, Zhang W, Ding Q. Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks. IEEE Transactions on Industrial Electronics. 2018, 66, 5525-34.

[18] Zhang A, Wang H, Li S, Cui Y, Liu Z, Yang G, et al. Transfer learning with deep recurrent neural networks for remaining useful life estimation. Applied Sciences. 2018, 8, 2416.

[19] Xu Y, Sun Y, Liu X, Zheng Y. A digital-twin-assisted fault diagnosis using deep transfer learning. IEEE Access. 2019, 7, 19990-9.

[20] Li C, Zhang S, Qin Y, Estupinan E. A systematic review of deep transfer learning for machinery fault diagnosis. Neurocomputing. 2020.

[21] Guo J, Wu J, Zhang S, Long J, Chen W, Cabrera D, et al. Generative Transfer Learning for Intelligent Fault Diagnosis of the Wind Turbine Gearbox. Sensors. 2020, 20, 1361.

[22] Shao J, Huang Z, Zhu J. Transfer Learning Method Based on Adversarial Domain Adaption for Bearing Fault Diagnosis. IEEE Access. 2020, 8, 119421-30.

[23] Zhang M, Wang D, Lu W, Yang J, Li Z, Liang B. A deep transfer model with wasserstein distance guided multi-adversarial networks for bearing fault diagnosis under

different working conditions. IEEE Access. 2019, 7, 65303-18.

[24] She D, Peng N, Jia M, Pecht M. Wasserstein distance based deep multi-feature adversarial transfer diagnosis approach under variable working conditions. Journal of Instrumentation. 2020, 15, 06002.

[25] Li X, Zhang W, Ma H, Luo Z, Li X. Deep learning-based adversarial multi-classifier optimization for cross-domain machinery fault diagnostics. Journal of Manufacturing Systems. 2020, 55, 334-47.

[26] Li J, Huang R, He G, Wang S, Li G, Li W. A Deep Adversarial Transfer Learning Network for Machinery Emerging Fault Detection. IEEE Sensors Journal. 2020.

[27] Jiao J, Zhao M, Lin J, Liang K. Residual joint adaptation adversarial network for intelligent transfer fault diagnosis. Mechanical Systems and Signal Processing. 2020, 145, 106962.

[28] Li X, Zhang W, Xu N-X, Ding Q. Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places. IEEE Transactions on Industrial Electronics. 2019, 67, 6785-94.

[29] Wang J, Chen Y, Hao S, Feng W, Shen Z. Balanced distribution adaptation for transfer learning. 2017 IEEE International Conference on Data Mining (ICDM). IEEE. 2017, 1129-34.

[30] Long M, Zhu H, Wang J, Jordan MI. Deep transfer learning with joint adaptation networks. International conference on machine learning. PMLR. 2017, 2208-17.

[31] Wang J, Feng W, Chen Y, Yu H, Huang M, Yu PS. Visual domain adaptation with manifold embedded distribution alignment. Proceedings of the 26th ACM international conference on Multimedia. 2018, 402-10.

[32] Yu C, Wang J, Chen Y, Huang M. Transfer learning with dynamic adversarial adaptation network. 2019 IEEE International Conference on Data Mining (ICDM). IEEE. 2019, 778-86.

[33] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image

recognition. arXiv preprint arXiv. 2014.

[34] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015, 1-9.

[35] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, 770-8.

[36] Zhang W, Peng G, Li C, Chen Y, Zhang Z. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors. 2017, 17, 425.

[37] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition. Ieee. 2009, 248-55.

[38] Ben-David S, Blitzer J, Crammer K, Pereira F. Analysis of representations for domain adaptation. Advances in neural information processing systems. 2006, 19, 137-44.

[39] Lessmeier C, Kimotho JK, Zimmer D, Sextro W. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. Proceedings of the European conference of the prognostics and health management society. 2016, 05-8.

[40] Li Y, Wang N, Shi J, Liu J, Hou X. Revisiting batch normalization for practical domain adaptation. arXiv preprint arXiv. 2016, 1603, 04779.

[41] Long M, Cao Y, Wang J, Jordan M. Learning transferable features with deep adaptation networks. International conference on machine learning. PMLR. 2015, 97-105.

[42] Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, et al. Domain-adversarial training of neural networks. The Journal of Machine Learning

Research. 2016, 17, 2096-30.

[43] Long M, Cao Z, Wang J, Jordan MI. Conditional adversarial domain adaptation. Advances in neural information processing systems. 2018, 1640-50.

[44] Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, et al. Automatic differentiation in pytorch. 2017.

[45] Sun B, Saenko K. Deep coral: Correlation alignment for deep domain adaptation. European conference on computer vision. Springer. 2016, 443-50.

[46] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation. International conference on machine learning. PMLR. 2015, 1180-9.

[47] Wang J, Zheng VW, Chen Y, Huang M. Deep transfer learning for cross-domain activity recognition. proceedings of the 3rd International Conference on Crowd Science and Engineering. 2018, 1-8.

**List of Tables**



**List of Figures**

Table 1. Conditions set in the fault diagnosis experiment for the wind turbine gearbox.

| Pattern No. | Fault location | Degree of failure | Label |
|---|---|---|---|
| **C0** | Ring gear | Missing tooth | 1 |
| **C1** | Ring gear | Crack tooth | 2 |
| **C2** | Sun gear | Missing tooth | 3 |
| **C3** | Planetary gear | Crack tooth | 4 |
| **C4** | Planetary gear | Missing tooth | 5 |
| **C5** | Normal | Normal (no failure) | 6 |

Table 2. Network structure parameters of the improved feature extractor in DTAL.

| Layers | Category | Parameters | Activation |
|---|---|---|---|
| 1 | Conv1d | OC=16, KS=64, S=1, PAD=32 | RL |
| | MaxPool1d | KS=2, S=2, PAD=0 | |
| 2 | Conv1d | OC=32, KS=3, S=1, PAD=1 | RL |
| | MaxPool1d | KS=2, S=2, PAD=0 | |
| 3 | Conv1d | OC=64, KS=3, S=1, PAD=1 | RL |
| | MaxPool1d | KS=2, S=2, PAD=0 | |
| 4 | Conv1d | OC=64, KS=3, S=1, PAD=1 | RL |
| | MaxPool1d | KS=2, S=2, PAD=0 | |
| 5 | Conv1d | OC=64, KS=3, S=1, PAD=1 | RL |
| | MaxPool1d | KS=2, S=2, PAD=0 | |
| 6 | Conv1d | OC=128, KS=3, S=1, PAD=1 | RL |
| | MaxPool1d | KS=2, S=2, PAD=0 | |
| 7 | linear | IF=2048, OF=256 | RL |

Remarks: OC- out_channels; KS- kernel_size; S- stride; PAD- padding; OS- output_size; IF- in_features; OF- out_features; and RL- ReLU.

Table 3. Network structure parameters of the label classifier in DTAL.

| Layers | Category | Parameters | Activation |
|---|---|---|---|
| Source_fc | linear | IF=256, OF=number of dataset categories | SM |

Remarks: IF- in_features; OF- out_features; and SM- SoftMax.

Table 4. Network structure parameters of the global domain discriminator and local domain discriminator in DTAL.

| Layers | Category | Parameters | Activation |
|---|---|---|---|
| layer1 | Linear | IF=256, OF=1024 | RL |
| | Dropout | p=0.5 | |
| layer2 | Linear | IF=1024, OF=1024 | RL |
| | Dropout | p=0.5 | |
| layer3 | Linear | IF=1024, OF=2 | SM |

Remarks: IF- in_features; OF- out_features; RL- ReLU; and SM- SoftMax.

Table 5. Fault diagnosis accuracies of different methods for the wind turbine gearbox experiment (%)

| **Transfer tasks** | 0-1 | 0-2 | 1-0 | 1-0 | 2-0 | 1-0 | Mean |
|---|---|---|---|---|---|---|---|
| **DTAL** | **70.14** | **66.01** | **69.93** | **76.94** | **71.56** | **79.76** | **72.39** |
| **DTAL-SFE** | 68.09 | 62.5 | 67.74 | 69.97 | 55.76 | 62.88 | 64.49 |
| **SCNN** | 48.11 | 50.4 | 53.91 | 65.52 | 47.13 | 63.84 | 54.82 |
| **AdaBN** | 56.36 | 53.28 | 53.86 | 69.67 | 50.57 | 56.98 | 56.79 |
| **MK-MMD** | 46.22 | 50.56 | 58.65 | 69.76 | 63.26 | 78.37 | 61.13 |
| **JMMD** | 48.37 | 49.41 | 59.86 | 73.09 | 70.8 | 79.44 | 63.5 |
| **DANN** | 42.12 | 42.99 | 60.24 | 71.25 | 67.5 | 77.71 | 60.3 |
| **CDAN** | 50.94 | 47.92 | 62.88 | 72.99 | 68.44 | 78.3 | 63.58 |

Table 6. Fault diagnosis accuracies of different methods for the benchmark bearing experiment (%)

| Transfer tasks | 0-1 | 0-2 | 0-3 | 1-0 | 1-2 | 1-3 | 2-0 | 2-1 | 2-3 | 3-0 | 3-1 | 3-2 | Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DTAL** | **67.08** | **97.07** | **85.48** | **83.03** | **84.68** | **63.14** | **96.25** | **70.26** | **86.25** | **83.25** | **52.31** | **89.39** | **79.85** |
| **DTAL-SFE** | **46.35** | **90.75** | **57.45** | **43.75** | 44.94 | 34.07 | **88.56** | **43.46** | **60.79** | **54.9** | **34.6** | 56.73 | **54.7** |
| **SCNN** | 14.52 | 88.63 | 35.39 | 27.18 | 29.87 | 17.91 | 87.14 | 14.32 | 43.93 | 36.87 | 23.53 | 36.06 | 37.95 |
| **AdaBN** | 19.06 | 89.24 | 41.79 | 28.02 | 33.57 | 23.21 | 86.86 | 17.88 | 46.05 | 40.36 | 26.27 | 39.2 | 40.96 |
| **MK-MMD** | 34.21 | 83.93 | 40.24 | 40.82 | 39.55 | 28.81 | 82.05 | 36.51 | 40.82 | 50.43 | 26.3 | 53.57 | 46.44 |
| **JMMD** | 38.86 | 85.19 | 47 | 41.12 | 44.23 | 29.29 | 82.15 | 39.15 | 46.49 | 54.7 | 30.45 | 57.44 | 49.67 |
| **DANN** | 35.96 | 84.05 | 39.71 | 41.75 | 36.84 | 32.24 | 81.43 | 40.67 | 42.28 | 51.84 | 30.98 | 54.41 | 47.68 |
| **CDAN** | 36.07 | 85.61 | 51.86 | 41.38 | **45.16** | **35.43** | 83.54 | 38.91 | 49.98 | 54.05 | 32.07 | **58.14** | 51.02 |

Table 7. Fault diagnosis accuracies of different methods (all with the improved feature extractor) for the wind turbine gearbox experiment (%)

| Transfer tasks | 0-1 | 0-2 | 1-0 | 1-0 | 2-0 | 1-0 | Mean |
|---|---|---|---|---|---|---|---|
| **DTAL** | **70.14** | **66.01** | **69.93** | **76.94** | **71.56** | 79.76 | **72.39** |
| **SCNN-IFE** | 59.56 | 52.94 | 62.63 | 73.69 | 55.91 | 80.74 | 64.24 |
| **AdaBN-IFE** | 63.49 | 57.43 | 63.98 | 73.28 | 62.26 | 80.41 | 66.81 |
| **MK-MMD-IFE** | 51.32 | 55.66 | 62.12 | 74.86 | 63.99 | 80.73 | 64.78 |
| **JMMD-IFE** | 53.02 | 59.83 | 61.7 | 72.05 | 64.37 | 78.09 | 64.84 |
| **DANN-IFE** | 59.2 | 63.89 | 66.39 | 74.13 | 67.12 | 79.51 | 68.37 |
| **CDAN-IFE** | 59.48 | 64.86 | 65.31 | 76.04 | 66.22 | **82.57** | 69.08 |

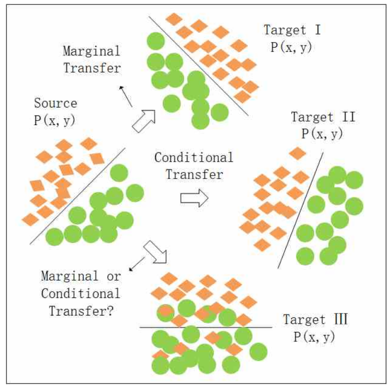


Fig. 1. Different influences of marginal and conditional distributions on transfer learning, where $P(x, y)$ denotes the distribution of the two domains.

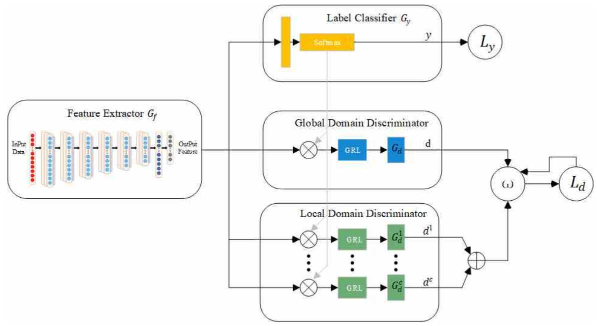


Fig. 2. Adversarial principle of the present DTAL.

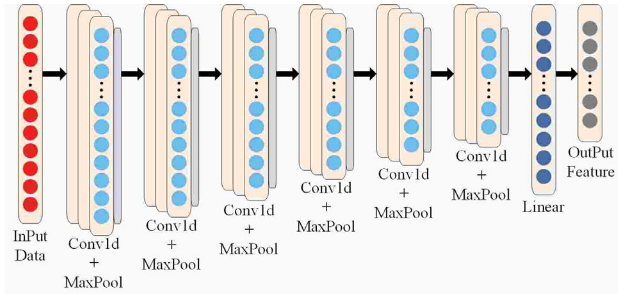


Fig. 3. The structure of the improved feature extractor.

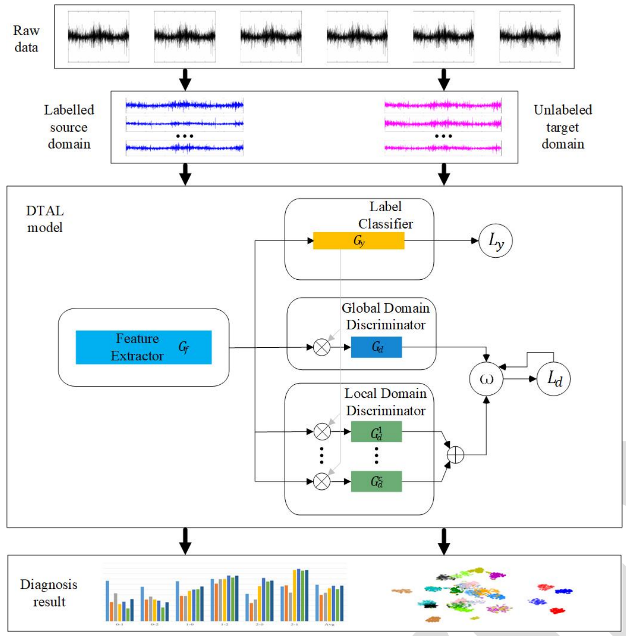


Fig. 4. Schematic of the proposed method.

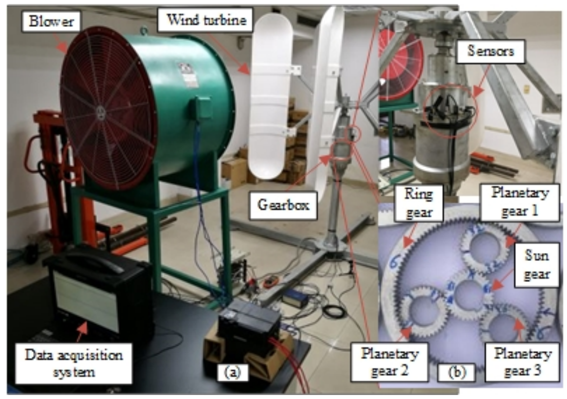

Fig. 5. Wind turbine experimental setup.

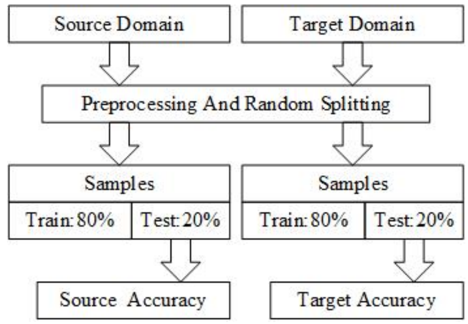


Fig. 6. Data splitting for all diagnosis scenarios.

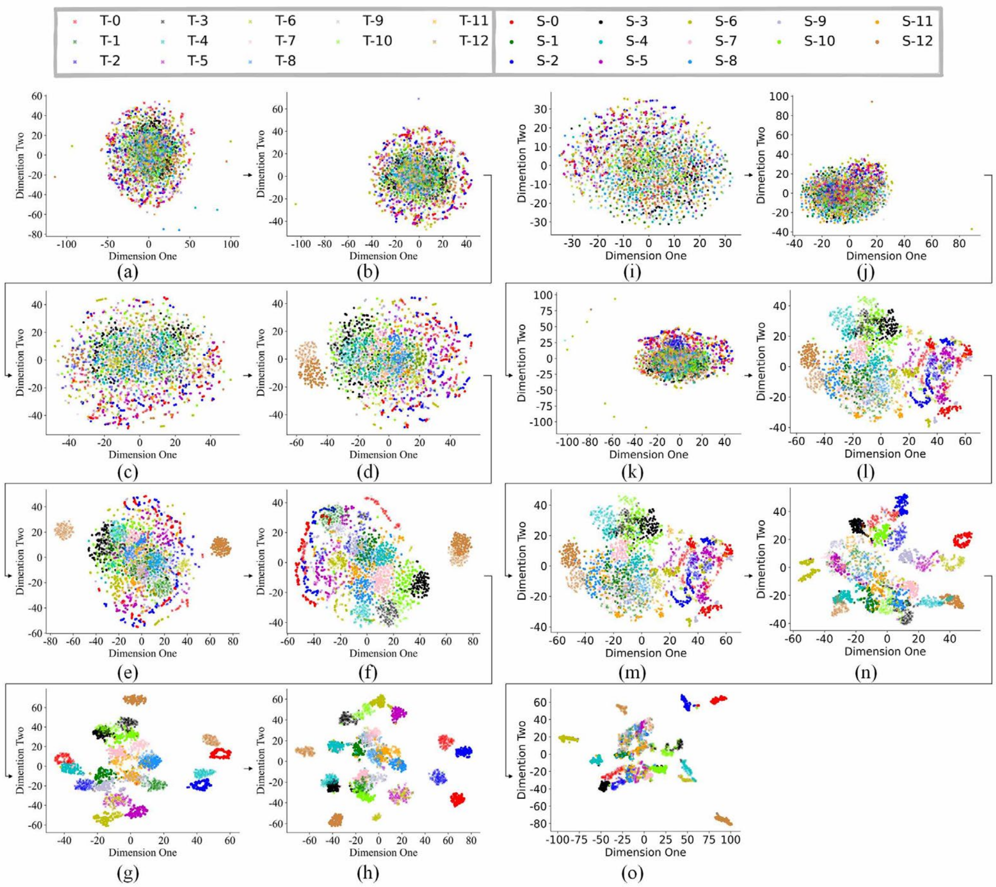


Fig. 7. Two-dimensional clustering diagrams of the data output of the DTAL: (a)-(h) with the improved feature extraction layer; and (i)-(o) with the original feature extraction layer.