This article has been accepted for publication in IEEE Transactions on Industrial Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3217541

Unbalanced Bearing Fault Diagnosis under Various Speeds based on Spectrum Alignment and Deep Transfer Convolution Neural Network

Feiyu Lu, Qingbin Tong, Ziwei Feng and Qingzhu Wan

***Abstract*—Bearing fault diagnosis plays a pivotal role in the safe and reliable operation of modern mechanical systems. However, the existing fault diagnosis methods rarely deal with the problem of category imbalance and various speeds concurrently, which cannot work effectively in practical scenarios. Considering the underlying similarities of data in frequency domain, data mining under various speeds can help to reduce the deviation of domain distribution. Therefore, a novel fault diagnosis method based on spectrum alignment (SA) and deep transfer convolution neural network (DTCNN) is proposed, where the SA and data augmentation module (DAM) are designed to extract spectrum alignment features from the unbalanced bearing data. The DTCNN model based on joint distribution adaptation (JDA) is built to facilitate learning reliable domain-invariant features. Different from the existing studies, a more general transfer task with time-varying speed is considered, even with complex faults. For 14 transfer tasks in two unbalanced fault diagnosis cases under variable speed, the average accuracy, F1-score, area under curve (AUC) of the proposed method can reach more than 97.76% 97.57% 98.75%. The results show that this method has superior diagnostic effect and better generalization ability than various state-of-the-art methods.**

***Index Terms*—Rolling bearing, convolution neural network, various speeds, spectrum alignment, unbalanced fault diagnosis**

I. Introduction

BEARING health monitoring and fault diagnosis play an important role in ensuring the reliability and safety of modern mechanical equipment. As an important component of rotating machinery, the failure of bearing will cause huge economic losses and even threaten personal safety. Therefore, the reliability and safety of bearing operations and better and more efficient bearing fault diagnosis technology have always been pursued.

This research was funded by the Beijing Natural Science Foundation (Grant no. L211010, 3212032), and the Fundamental Research Funds for the Central Universities (2022YJS155). (corresponding author: Qingbin Tong).

Feiyu Lu, Qingbin Tong and Ziwei Feng are with School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China. (e-mail: 21117039@bjtu.edu.cn; qbtong@bjtu.edu.cn;21121421@bjtu.edu.cn).

Qingzhu Wan is with School of Electrical and Control Engineering, North China University of Technology, Beijing 100144, China.

Deep learning (DL) [1] has obvious advantages in big data processing, and it has achieved great success in computer vision, speech recognition, image segmentation and other fields. At the same time, there are new developments in intelligent rolling bearing fault diagnosis [2]. DL can automatically extract features from a large number of vibration signals to achieve fault classification, which is not possible with traditional fault diagnosis methods in today's big data era. In recent years, many intelligent fault diagnosis models based on neural networks have been proposed and have achieved better performance. Deep learning methods have a large demand for samples with labels [3]. The trained deep learning models are highly accurate on test bearing datasets with the same data distribution. For example, Verstraete et al. [4] converted bearing vibration signals into time-frequency images and fed them into a deep convolutional neural network (DCNN) to achieve adaptive feature extraction. The experimental results show that the deep learning model can achieve better results with fewer parameters. Zhao et al. [5] proposed a bearing fault type identification method based on a multilayer extreme learning machine (ELM) without the feature extraction step in traditional fault diagnosis methods, which was validated in a publicly available bearing dataset. In addition, other deep learning models, such as autoencoders (AEs) and its variants [6, 7], deep belief network (DBN) [8], and deep temporal dictionary learning (DTDL) [9], have solved the feature extraction problem in different scenarios. However, these diagnostic models usually work only if the training and test sets are derived from the same speed conditions. In practical engineering applications, there is a wide variety of rotational speed conditions, and in most cases, the test set is untagged data.

To this end, transfer learning models have been proposed to address the above unsupervised fault diagnosis problem. Zhao et al. [10] described the principles of domain adaptation and its application in several research areas. To implement diagnosis under different operating conditions while taking into account the noise factor in the actual operating conditions, Zhang et al. [11] analyzed the relationship between dropout, batch normalization, and noise and proposed a convolutional neural network with training interference (TICNN) model, which was experimented on the Case Western Reserve University (CWRU) bearing dataset and achieved good diagnostic results. Wen et al. [12] exploited a deep transfer learning model based on a sparse autoencoder and maximum mean discrepancy (MMD), and a comparative validation experiment was also conducted on the CWRU dataset. In reference [13], to solve the problem that it is

difficult to obtain the label in the target domain of the bearing under variable working conditions, adaptive logarithm normalization (ALN) is proposed to preprocess the data, and then ALN and simplified lightweight models are integrated into a deep adaptation network (DAN), which was validated in two publicly available bearing datasets.

On the other hand, in addition to the aforementioned domain adaptation problem, sample imbalance is also a challenge for fault diagnosis [14]. Bearing sample data collected in real industrial environments are often unbalanced, i.e., there are enough normal samples and not enough fault samples, and there is variability between the number of each fault. For brevity, only a few representative works are presented. Jia et al.[15] proposed the deep normalized convolutional neural network (DNCNN) to solve the question of fault classification under unbalanced bearing data and explored the interpretability of the proposed model using the neuronal activation maximum algorithm. Pan et al. [16] used GAN to generate fault samples that did not exist originally; thus, the sample data were balanced, and a comparative validation experiment was performed on three bearing datasets. Khodayar et al. [17] developed a convolutional graph autoencoder model that can generate samples from the probability density learned by each node. Zhao et al. [7] proposed a variational autoencoder-based data augmentation technique and verified that the generated data and the real data had similar fault characteristics to each other through a time-frequency analysis method.

From a review of the literature in the above paragraph, most intelligent diagnostic methods are dedicated to a single problem in variable speed or imbalance. To the best of our knowledge, research on the unbalanced classification of bearings with different speeds is still quite rare. The data distribution under time-varying speed may change with time, that is, there is conceptual drift [18]. Concept drift can conduct online learning on dynamic data flow to adapt to changing data distribution [19]. The experimental section below presents the results of the comparative methods based on concept drift detection. Xu et al. [20] used three loss functions, including label loss, domain loss and contrast loss, to integrate a fusion fault diagnosis network called the renewable fusion fault diagnosis network (RFFDN) to achieve the unification of variable speed and unbalanced problems; however, the method did not take into account the time-varying speed conditions, and the unbalance of the fault samples was low. Wang et al. [21] proposed a novel adaptive normalized convolutional neural network (ANCNN) combined with the Teager Calculated Order (TCO) spectrum to solve the problem of intelligent fault diagnosis in time-varying speed conditions of gearboxes. However, the method is computationally intensive due to the existence of iterations, while its application within the field of bearing diagnosis has yet to be explored.

This paper proposes an intelligent fault diagnosis model based on SA and DTCNN, which can deal with the time-varying speed fault diagnosis problem under unbalanced samples as well as the constant speed fault classification task. The proposed model has a strong generalization capability. The main contributions of this paper are summarized as follows.

(1) We propose an SA algorithm that can convert nonstationary vibration signals into stationary signals, use SA to extract valid fault features from time-varying rotational speeds, and evaluate the necessity of the method for diagnostic models.

(2) A data augmentation module (DAM) combining density-based spatial clustering of applications with noise (DBSCAN) and an unbalanced sampler is introduced to address the problem of overfitting of most classes in unbalanced samples.

(3) We try to integrate the classification tasks under time-varying speed and constant speed in the proposed SA-DTCNN method, which achieves the unification of the model and improves the generalization ability of the intelligent diagnosis model in the face of changeable speed. For the first time, we have tried and achieved the unification of the problems of constant and time-varying speed intelligent fault diagnosis under unbalanced samples and domain shifts.

The remainder of the article is structured as follows: In Section 2, we give the problem definition. In Section 3, the proposed fault classification framework is presented. Details of two different rolling bearing datasets and the corresponding experimental results are presented in Section 4. Section 5 is the conclusion.

II. Problem Definition

We study the unbalanced fault diagnosis of bearings under variable speed conditions. Specifically, we take the training data as the source domain data and the test data as the target domain data. Datasets D^m under various speeds, where m means the number of speeds. The source domain $D_S^{m\_s} = \{(\boldsymbol{x}_s^{(i)}, \boldsymbol{y}_s^{(i)})\}_{i=1}^N$ and the target domain $D_T^{m\_t} = \{(\mathbf{x}_t^{(i)})\}_{i=1}^M$ respectively correspond to the bearing vibration signal at different speeds, and $D^m \supseteq \{D_S^{m\_s} \bigcup D_T^{m\_t}\}$, N, M are the number of samples in the source domain and the target domain, respectively, m_s, m_t are the corresponding speed. In addition, the labels of $D_S^{m\_s} = \{(\boldsymbol{x}_s^{(i)})\}_{i=1}^N \{(\boldsymbol{y}_s^{(i)})\}_{i=1}^N$ are known, and the labels of $D_T^{m\_t} = \{(\boldsymbol{x}_t^{(i)})\}_{i=1}^M$ are unknown. Let P, Q be the feature discrepancy of $D_S^{m\_s}$ and $D_T^{m\_t}$. Usually, due to domain shift, $P \neq Q$. Moreover, the number of labels in the source domain sample and the target domain sample is unbalanced.

Based on the above description, we take normal and two fault labels as examples. Fig. 1 shows the three challenges mentioned. In the following experimental stage, the data we use are variable speed and unbalanced, and this fault diagnosis mode is used to simulate the real industrial fault scene as much as possible.

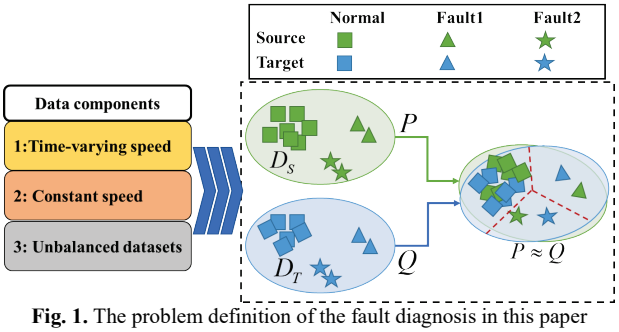


Fig. 1. The problem definition of the fault diagnosis in this paper

III. The Proposed Fault Classification Framework

A. Spectral Alignment (SA)

The frequency domain research method is a commonly used signal analysis method, which allows time series to be transferred into Fourier space and equated to frequency spectral lines, where the corresponding frequency values of the spectral lines contain the frequency information of the signal. Without considering noise, the fault impulse response in a rolling bearing vibration signal can be seen as a composition of spectral lines arranged according to certain rules [22]. The outer ring, inner ring, rolling element fault characteristic frequency and rotational frequency of a rolling bearing can be used as a bridge to represent the rules between these spectral lines, as shown in Eq. (1).

$$\begin{cases} f_0 = \frac{1}{2} Z(1-\frac{d}{D}\cos\alpha) f_r = \delta_r f_r \\ f_i = \frac{1}{2} Z(1+\frac{d}{D}\cos\alpha) f_r = \delta_i f \\ f_b = \frac{D}{2d}[1-(\frac{d}{D})^2 \cos^2\alpha] f_r = \delta_b f \end{cases} \tag{1}$$

where D is the bearing pitch diameter; d is the rolling element diameter; f_r is the rotation frequency; f_0, f_i, f_b represent the outer ring, inner ring, and rolling element fault characteristic frequencies, respectively; $\delta_o, \delta_i, \delta_b$ are the fault characteristic coefficients; and α is the contact angle between the rolling element and the raceway.

From Eq. (1), we can see that once f_r changes, f_0, f_i, f_b will change, i.e., the spectrum changes. In the constant speed condition, we assume that given speed *a* and speed *b*, where *a* and *b* are constant values, the vibration signal dataset A and dataset B are obtained from the acquisition instrument. From Eq. (1), it can be concluded that the spectral lines corresponding to each sample in dataset A or B are the same and have the same distribution law. For the transfer task from speed *a* to speed *b*, it is only necessary to learn the distribution laws of the two different spectral lines, which is not difficult, and there have been many successful cases [23, 24]. However, in the case of time-varying speed conditions, which are common in aviation and transportation, if we were to perform the transfer task from dataset A to dataset B, where the rotational speeds within A and B are transformed values, we would face the problem that the distribution of spectral lines would be different for each sample, which would lead to the failure of the current state-of-the-art algorithm.

From the above analysis, it can be seen that if the vibration signal under time-varying speed can be transformed into a constant speed vibration signal, we can use the current intelligent diagnosis model for classification. To this end, this paper develops an SA algorithm based on generalized demodulation (GD) [25]. Its main core idea is to eliminate the influence of speed fluctuation in vibration signal and then align the spectrum of all samples based on the first sample.

First, the principle of the GD algorithm is as follows.

$$X_G(f) = \int_{-\infty}^{\infty} x(t)\exp(-2\pi j(ft + s(t)))\mathrm{d}t \tag{2}$$

where $x(t)$ is the vibration signal, f_s is the sampling frequency, $s(t)$ is called the generalized demodulation operator (GDO) and the calculation equation is as follows.

$$s(t) = \int (\frac{\mathrm{d}\varphi(t)}{2\pi \mathrm{d}t} - f_0)\mathrm{d}t \tag{3}$$

where $\varphi(t)$ is the phase function and f_0 represents the initial speed frequency. Both of them can be obtained from the bearing speed signal $p(t)$.

Then, the GD algorithm and filter bank are used to eliminate the speed fluctuations and obtain a smooth vibration signal, that is, a spectrum in which the spectral lines are perpendicular to the frequency axis. The equation is as follows.

$$X = X_G^i(f), \qquad f_{a(n)}(i) \le f \le f_{b(n)}(i) \tag{4}$$

where i is the number of filters, $i = 1, 2, \cdots 11$, $a(n)$, and $b(n)$ corresponds to the n-th sample.

$$\begin{cases} f_{a(n)}(i) = kf_r - \Omega_2 \\ f_{b(n)}(i) = kf_r - \Omega_2 \end{cases} \tag{5}$$

where $\mathrm{k} = 1, 2, \delta_o, \delta_i, \delta_b, 2\delta_o, 2\delta_i, 2\delta_b, 3\delta_o, 3\delta_i, 3\delta_b$, and Ω_1, Ω_2 are the filter parameters.

Finally, the spectrum corresponding to all samples is aligned based on the first sample with the following equation.

$$X'^i_n\left[f_{a(0)}(i), f_{b(0)}(i)\right] = X^i_n\left[f_{a(n)}(i), f_{b(n)}(i)\right] \tag{6}$$

X^i_n is the spectral line corresponding to the n-th sample, and X'^i_n is the aligned spectral line. The pseudocode of SA is described in Algorithm 1.

Algorithm 1: Spectral Alignment (SA)

Input:

- Raw signal $x(t)$, speed function f_r, number of samples N
- Initial $X'^i_n = 0$, $X' \in [1, f_s/2]$
- $\delta_o, \delta_i, \delta_b$, Ω_1, Ω_2, $\mathrm{k} = 1, 2, \delta_o, \delta_i, \delta_b, 2\delta_o, 2\delta_i, 2\delta_b, 3\delta_o, 3\delta_i, 3\delta_b$

Output: spectrum after alignment X'^i_n

for n from 1 to N **do**
  for i from 1 to 11 **do**
    $f_{a(n)}(i), f_{b(n)}(i) \leftarrow f_r$ # Filter parameters using Eq. (5)
    $s(t) \leftarrow \int (f_r(t) - f_0)\mathrm{d}t$ # Parameter for GD (Eq. (3))
    $X^i_n \leftarrow x(t), s(t)$ # GD using Eq. (2)
    $X'^i_n \leftarrow X^i_n\left[f_{a(n)}(i), f_{b(n)}(i)\right]$ #Eq. (6)
  end for
end for

B. Data augmentation module (DAM)

The problem of unbalanced data has received increasing attention [14]. In the case of unbalanced data samples, the performance of intelligent fault diagnosis models will be greatly reduced.

To reduce the dominant effect of most categories on the model and improve the fitting effect of a few categories in unbalanced samples, we develop a data augmentation module, which is composed of two parts. First, the feature data are classified based on the DBSCAN clustering method [26]. This process only needs to evaluate the number of each category, and it is acceptable without labels. DBSCAN is easy to implement in machine learning toolkits. Second, we randomly select samples from each cluster of data with $p_{label} = 1/N_l$ to realize the equilibrium of categories, where N_l h is the number of each category. This process can be realized by an unbalanced sampler. The imbalanced data sampling is not simply oversampling [27] or undersampling [28] but resampling the imbalanced data according to the weight of the labels, which avoids overfitting to some extent. IDS was validated in the imbalanced MNIST dataset. Reference code link [1]. The pseudocode of DAM is described in Algorithm 2.

Algorithm 2: Data Augmentation Module (DAM)

Input:

- Dataset D which containing L labels.
- MinPts: the threshold of field density.
- Radius ε.

Output: The balanced dataset D_b

1: Mark all samples in D as unvisited data
2: **Do**
3: Randomly select a sample p from the unvisited data. M is the number of samples in field p-ε. ϕ is a collection within the field p-ε.
4: **if** $M \geqslant$ MinPts
5: Create l-th cluster C_l, and put p into C_l.
6: **for** p' in ϕ **do**
7: **if** p' is unvisited
8: Mark p' as a visited data.
9: **if** the number of p'-$\varepsilon \geqslant$ MinPts
10: Put the samples in the field p'-ε into ϕ.
11: **if** $p' \notin C$
12: Put p' into C_l.
13: **end for**
14: Output C_l
15: **else** p is noise
16: until there are no unvisited samples in D.
17: N_l is the number of label l in C_l
18: $D_b \leftarrow D$ # Resample with a probability of $p_{label} = 1/N_l$

C. Deep transfer convolutional neural networks

In recent years, convolutional neural networks (CNNs) have received a great deal of attention from scholars due to their powerful feature extraction capabilities. The proposed DTCNN model consists of five convolution layers, five pooling layers and two fully connected layers (FCs). BN technology is embedded in each convolution layer. ReLU is used as the activation function for the convolution of layers 1-4, while the activation function for the convolution of layer 5 is RReLU, followed by the dropout layer.

Two loss functions, the cross entropy loss function and the transfer regular term function, are used. In multiclassification tasks, the cross entropy loss function is as follows.

$$\mathcal{L}_c = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C}\mathbf{y}_c^{(n)}\log\hat{\mathbf{y}}_c^{(n)} \tag{7}$$

where N is the number of samples, C is the label type, $\mathbf{y}_c^{(n)} \in \{0,1\}$ is a sign function, and $\hat{\mathbf{y}}_c^{(n)}$ is the probability that the n-th sample is predicted to be C.

In the domain adaptation task, $P(X_s, Y_s)$ represents the joint distribution of features X_s and labels Y_s in the source domain, and $Q(X_t, Y_t)$ represents the joint distribution in the target domain. To bring these two joint distributions as close as possible for better unsupervised fault diagnosis, the JMMD loss function is introduced as the transfer regular term. Its equation is as follows [29].

$$\mathcal{L}_{\mathrm{JMMD}}(P,Q) = \left\| \mathbb{E}_P\left[\otimes_{l\in L}\phi\left(\mathbf{X}^{sl}\right)\otimes\psi\left(\mathbf{Y}^{s}\right)\right] - \mathbb{E}_Q\left[\otimes_{l\in L}\phi\left(\mathbf{X}^{tl}\right)\otimes\psi\left(\mathbf{Y}^{t}\right)\right]\right\|_{\mathcal{H}}^2 \tag{8}$$

where $\mathbb{E}$ is the mathematical expectation and $\mathcal{H} \triangleq \otimes_{l\in L}\mathcal{H}_\phi \otimes \mathcal{H}_\psi$ is the tensor product of $\mathbf{X}$ and $\mathbf{Y}$ of Hilbert spaces. L is the number of layers of the network. We adopt the features of a network layer, so the value of L is 1. According to the design idea in reference [30], the JMMD has four input parameters, which are the output vectors of FC1 and FC2, $\mathbf{X}^s$, $\mathbf{Y}^s$, $\mathbf{X}^t$, and $\mathbf{Y}^t$.

The total loss function can be written as

$$\mathcal{L} = \mathcal{L}_c + \lambda_{\mathrm{JMMD}}\mathcal{L}_{\mathrm{JMMD}}\left(P_s, D_t\right) \tag{9}$$

where λ_{JMMD} is the weight coefficient with a variable value, and its calculation equation is as follows.

$$\lambda_{\mathrm{JMMD}} = \frac{2}{1+\exp(-10\times q/Q)} - 1 \tag{10}$$

where q is the number of model training iterations and Q is the total number of training iterations.

D. Framework structure

When a rolling bearing fails, the fault shock response in the spectrogram can be interpreted as a fault signature, which is a valid feature representation. However, the fault features in the spectrograms at constant speeds and time-varying speeds are different, which makes the diagnosis methods in the two cases inconsistent. To unify the fault diagnosis models under the two scenarios and improve the diagnostic accuracy, we use the proposed SA algorithm to preprocess the original vibration signal, from which effective fault features are extracted, and then use the DTCNN model for the fault classification task. The overall SA-DTCNN fault diagnosis model is shown in Fig. 2.

Combined with Fig. 3, the overall process of the proposed method for classifying unbalanced fault samples of bearings under variable speed conditions in this paper is as follows:

[1] https://github.com/ufoym/imbalanced-dataset-sampler

(1) The original signals of rolling bearings under different speed conditions are slidingly partitioned into the required training and test sets in fixed strides.

(2) The SA algorithm is used to process the dataset to obtain a one-dimensional feature vector. Subsequently, the DAM is utilized to process the imbalanced dataset.

(3) Start training the model. The feature values after SA and DAM are used as input to the DTCNN model, and the backpropagation algorithm is executed to optimize the parameters of the model by minimizing the value of the loss function in Eq. (16). It is worth stating that training tricks are added in this paper with reference to [31]. A so-called pretraining session with an epoch of 20 is used, as well as an early stop mechanism at epoch 25, with a maximum training epoch of 500.

(4) Input the data of the testing data (target domain) into the trained model, automatically identify the fault type and complete the classification task.

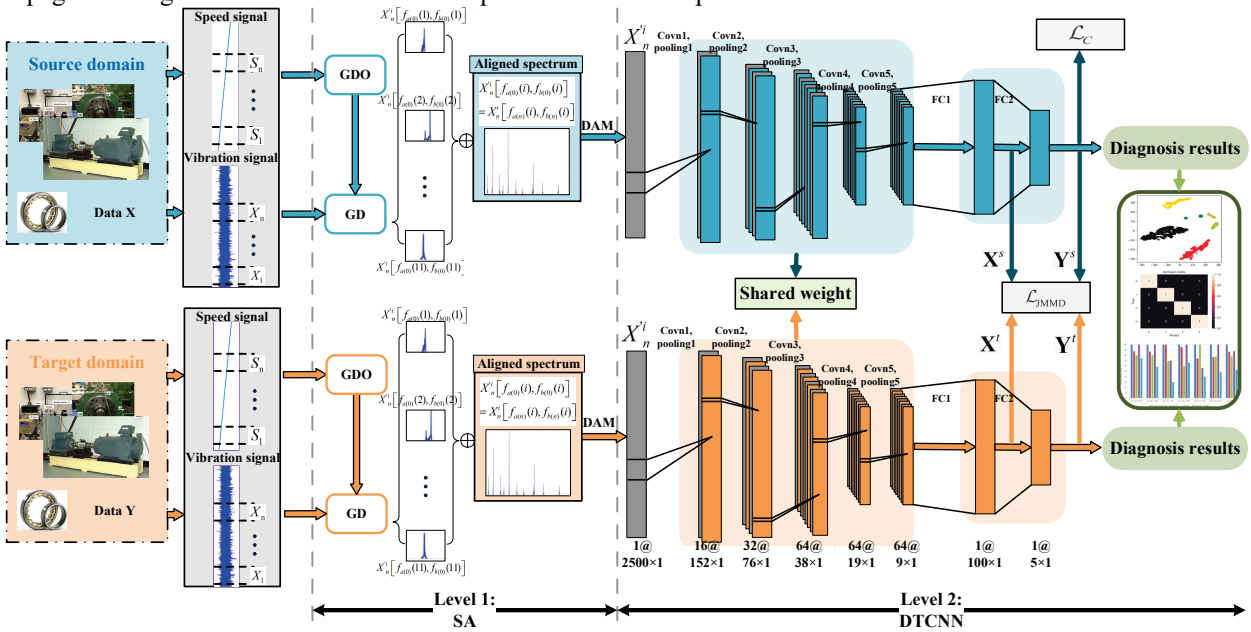


Fig. 2. Illustration of the proposed fault classification method

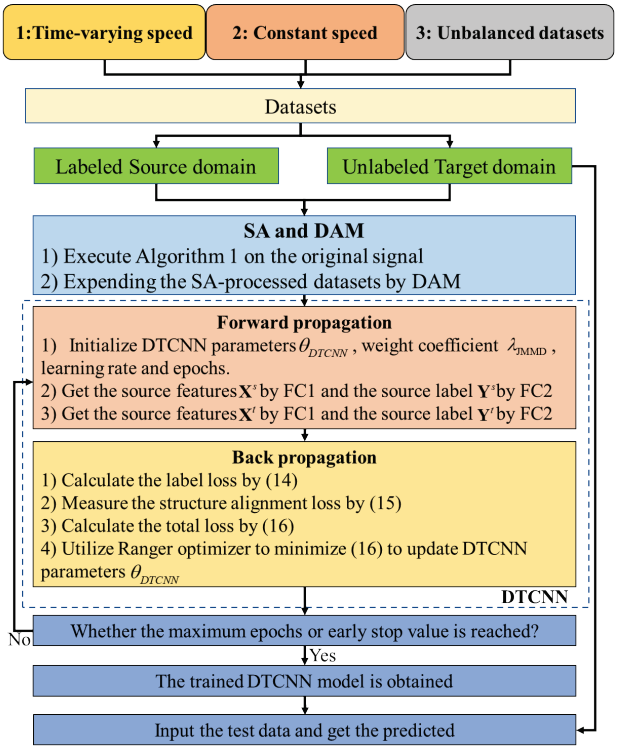


Fig. 3. Flowchart of the proposed fault classification method

IV. Experimental Validation

A. Dataset description

(1) Canada datasets: The Machinery fault simulator experimental platform (MFS-PK2 M) [32] is controlled by an AC drive that controls the motor and thus turns the rotating shaft. The accelerometer is installed in the 12 o'clock direction of the experimental bearing to obtain the raw vibration signal. A speed encoder (EPC model 775) is installed to monitor the speed of the bearings, and the cycles per revolution is 1024. The vibration and speed signals corresponding to the different health conditions of the bearings are tested at different speed fluctuations with a sampling frequency of 200 kHz and a sampling time of 10 s. The dataset contains five types of bearing health statuses, including normal condition (NC), outer race faults (OF), inner race faults (IF), ball faults (BF) and compound faults (CF) on the outer race, the inner race and a ball. There are 500 samples for each health condition, with a time length of 5 s for each sample and a step size of 0.01 s. Because there are five fault states, a total of 2500 samples are obtained. The velocity function corresponding to each sample is calculated based on the pulse sequence method.

(2) CWRU datasets: CWRU datasets [33] that were collected under constant speed are utilized for analysis. The sampling frequency at the driving end is 12 kHz. There are four health states: NC, BF, IF and OF. In addition to NC, each fault type contains three different fault diameters, representing the degree of damage. In other words, there are 10 state types. The

data of three different working conditions are obtained at 1772 rpm, 1750 rpm and 1730 rpm. In each fault state, the length of each sample is 2.5 s, and 500 sample data are obtained by sliding in strides of 0.005 s. Because there are ten fault states, a total of 5000 samples are obtained.

B. Approaches for comparison and implementation details

According to the description in references [14, 34], we use three current mainstream imbalance identification technologies for comparison, namely, **synthetic samples**: SMOTE. **Re-weighting:** CB Loss, Focal Loss. **Model-based:** MLP, DNCNN, RFFDN, TICNN.

SMOTE (Method 1): SMOTE [35] is a classic work of data synthesis technology. SOMTE is implemented, where DAM was removed from the proposed method. CB Loss (Method 2): CB Loss is the category balanced softmax cross entropy loss function. Focal Loss (Method 3): Similarly, DAM was removed from the proposed method. SA+DAM+MLP (Method 4): Based on the proposed method, DTCNN was changed to MLP for the benchmark. SA+DAM+DNCNN (Method 5): DNCNN is a model-based unbalanced fault classification method, which has been proven to be effective in reference [15]. SA+DAM+RFFDN (Method 6): RFFDN [20] is an unbalanced classification method based on the combination of the model and loss function. SA+DAM+TICNN (Method 7): The fault diagnosis effect of TICNN under cross-speed conditions has been confirmed. The experimental setup details for each method are presented in Table I.

TABLE I
EXPERIMENTAL SETUP DETAILS FOR EACH METHOD

Method	Experimental setup details
SMOTE	Implemented by *imblearn.over_sampling* package, it is available in the PyPI.
CB Loss	Its parameter settings: beta=0.9999, gamma=2.0.
Focal Loss	Its parameter settings: alpha=1, gamma=1.
MLP	Two-layer fully connected neural network structure with input and output size of [2500-1200-600].
DNCNN	Two-layer CNNs, consistent with [15].
RFFDN	Five-layer CNNs, consistent with [20].
TICNN	Six-layer CNNs, consistent with [11].
Proposed	Five-layers CNNs, details are in Part C and D of Section 3.

Considering fairness, in each case, the hyperparameters used in the comparison method are consistent, and the experimental results are the average of five experiments. More importantly, in the process of model training and testing, to simulate the real industrial fault diagnosis scenario, the cross test follows the following rules: 1) For the source domain and the target domain, the data of the training set and the test set do not overlap. 2) The training set in the source domain and the test set in the target domain correspond to different speed ranges. In other words, the unbalanced training set in the source domain is used to train for the testing set in the target domain, and the unbalanced training set in the target domain is used to train for the testing set in the source domain. Each domain corresponds to data at different speeds. As shown in Fig.4. For each case, the required parameter settings are shown in Table II.

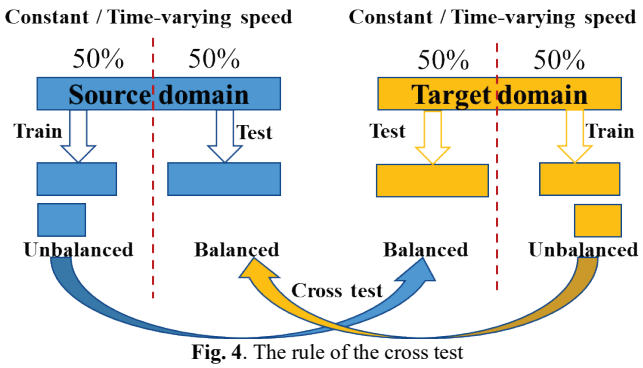


Fig. 4. The rule of the cross test

TABLE II
PARAMETER SETTINGS FOR EACH CASE

	[Ω1, Ω2]	[ε, MinPts]	Learning rate (Lr)	Batch size (Bs)
Case 1	[1, 2]	[3, 5]	0.002	25
Case 2	[3, 6]	[3, 5]	0.007	100

C. Evaluation metrics

To measure the effect of the proposed method on unbalanced fault classification across speeds, following previous studies [34], we use three quantitative indicators to measure the test effect of the model. They are the average fault identification accuracy (Acc), F1-score (F1) and average area under the receiver operating characteristic curve (AUC).

$$\mathrm{Acc}=\frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

$$\mathrm{F1}=\frac{2TP}{2TP+FP+FN} \tag{12}$$

where TP is the true positive (positive samples are successfully predicted as positive), TN is the true negative (negative samples are correctly predicted to be negative), FP represents the false positive (negative samples are incorrectly predicted to be positive), and FN is the false negative (positive samples are incorrectly predicted to be negative).

D. Case 1 Ottawa University datasets

1) Experiment setup

The vibration signals of the bearing at the ascending and descending speeds are collected for each health condition, corresponding to Dataset A and Dataset B.

To test the performance of the proposed method under unbalanced datasets, we use 50% of the data as the training set and the remaining part as the test set. The details of each dataset are summarized in Table III.

2) Results and analysis

In this section, we measure the capability of the proposed method in processing the vibration signals of time-varying speed from several perspectives and compare it with seven recent fault diagnosis methods.

After the feature extraction of the original vibration signal by SA, the proposed DTCNN model is used to perform fault classification on the dataset in Table IV. The method of intersection of speed up data and speed down data is adopted for testing. For example, the label of speed up dataset A is known, and the label of speed down dataset B is unknown. Datasets A1 and B1 are used as the training set, and B3 is used

as the test set. At this time, it is recorded as A1, B1 → B3. The test results are listed in Table 3. It can be seen from the table that the proposed method achieves diagnostic accuracy, F1, AUC of 99.23%, 99.20%, and 99.52%. Compared with Methods 1-7, the proposed method has the highest Acc, F1 and AUC in 8 fault diagnosis tasks.

TABLE III
THE DETAILS OF EACH DATASET FOR CASE 1

Health condition	Labels	Dataset A				Dataset B			
		Speed function (Hz)	Dataset A1	Dataset A2	Dataset A3	Speed function (Hz)	Dataset B1	Dataset B2	Dataset B3
NC	0	15.3+0.9×t	50%	50%	50%	28.95-1.5×t	50%	50%	50%
OF	1	15.69+1.2×t	40%	30%	50%	24.98-1.5×t	40%	30%	50%
IF	2	12.47+1.5×t	30%	20%	50%	24.26-1.48×t	30%	20%	50%
BF	3	14.1+1.05×t	20%	10%	50%	24.8-1.5×t	20%	10%	50%
CF	4	13.03 +1.49×t	10%	5%	50%	26.2 -1.5×t	10%	5%	50%

TABLE IV
CLASSIFICATION RESULTS ± STANDARD DEVIATIONS (%) BASED ON DIFFERENT FEATURES AND MODELS IN CASE 1

Cross test	Metrics	A1, B1 →B3	A1, B2 →B3	A2, B1 →B3	A2, B2 →B3	B1, A1 →A3	B1, A2 →A3	B2, A1 →A3	B2, A2 →A3	Average
Method 1	Acc	84.46±12.15	75.58±7.94	67.85±8.53	72.04±8.11	99.20±1.78	98.00±4.47	67.10±9.10	83.69±2.75	80.99
(SMOTE)	F1	82.06±14.82	71.14±9.62	63.15±11.51	65.68±10.15	99.19±1.80	98.10±4.23	62.61±11.16	81.56±4.48	77.93
	AUC	90.29±7.59	84.74±4.96	79.91±5.33	82.53±5.06	99.50±1.11	98.75±2.79	79.44±5.68	89.81±1.71	88.12
Method 2	Acc	85.31±5.38	63.36±6.97	68.55±8.42	61.36±5.89	98.11±2.71	99.90±0.21	81.21±11.36	89.53±7.48	80.91
(CB Loss)	F1	84.37±5.82	59.94±8.72	64.14±10.98	57.30±6.28	98.05±2.84	99.90±0.21	79.64±13.58	88.68±8.73	79.00
	AUC	90.82±3.36	77.10±4.35	80.35±5.26	75.85±3.68	98.82±1.69	99.94±0.13	88.25±7.10	93.46±4.67	88.07
Method 3	Acc	76.24±2.66	72.25±7.84	66.25±6.38	72.28±12.88	98.83±1.31	100.00±0.00	63.99±7.54	78.15±11.21	78.49
(Focal loss)	F1	71.51±3.50	66.60±9.90	59.81±7.30	67.62±17.10	98.81±1.33	100.00±0.00	57.94±8.71	74.80±13.97	74.63
	AUC	85.14±1.66	82.66±4.90	78.91±3.99	82.68±8.05	99.27±0.82	100.00±0.00	77.50±4.71	86.34±7.01	86.56
Method 4	Acc	84.97±11.28	78.78±11.07	81.76±5.90	87.20±5.55	63.69±7.01	68.43±7.03	73.76±15.56	71.07±16.99	76.20
(MLP)	F1	80.97±14.63	73.58±15.25	76.10±8.89	85.77±7.87	58.63±9.00	63.79±5.65	70.23±17.75	66.54±19.92	71.95
	AUC	90.61±7.05	86.74±6.92	88.60±3.69	92.00±3.46	77.31±4.38	80.27±4.39	83.60±9.73	81.92±10.62	85.13
Method 5	Acc	72.22±8.74	77.77±10.00	79.32±11.97	64.19±6.74	44.91±5.62	44.54±8.47	41.84±2.53	41.28±1.58	58.25
(DNCNN)	F1	69.41±9.74	75.53±10.94	78.10±12.99	60.31±8.76	40.37±7.68	37.47±9.70	37.26±5.19	37.60±5.67	54.50
	AUC	82.64±5.46	86.11±6.25	87.07±7.48	77.62±4.21	65.57±3.51	65.34±5.29	63.65±1.58	63.30±0.99	73.91
Method 6	Acc	74.73±11.42	78.43±13.74	81.80±18.02	74.65±23.18	78.97±21.15	80.28±25.56	89.07±23.54	64.03±21.31	77.74
(RFFDN)	F1	70.84±11.23	75.07±14.40	79.29±19.52	71.49±26.21	76.79±24.48	78.77±28.22	87.25±27.60	59.99±23.57	74.93
	AUC	84.21±7.14	86.52±8.59	88.63±11.26	84.16±14.48	86.86±13.22	87.68±15.97	93.17±14.71	77.52±13.32	86.09
Method 7	Acc	98.68±2.19	99.37±0.99	96.83±3.64	90.35±8.76	97.58±4.50	99.55±0.60	92.51±14.49	98.22±2.22	96.63
(TICNN)	F1	98.68±2.19	99.37±1.00	96.82±3.65	90.31±8.79	97.55±4.56	99.55±0.60	91.68±16.33	98.21±2.25	96.52
	AUC	99.18±1.37	99.61±0.62	98.02±2.27	93.97±5.47	98.49±2.81	99.72±0.37	95.32±9.06	98.88±1.39	97.89
Proposed	Acc	**99.96±0.04**	**95.32±5.53**	**99.58±0.88**	**99.18±1.04**	**99.98±0.03**	**99.96±0.07**	**99.88±0.20**	**100.00±0.00**	**99.23**
	F1	**99.96±0.04**	**95.13±5.83**	**99.58±0.88**	**99.18±1.04**	**99.98±0.03**	**99.96±0.07**	**99.88±0.20**	**100.00±0.00**	**99.20**
	AUC	**99.98±0.02**	**97.08±3.45**	**99.74±0.55**	**99.49±0.65**	**99.99±0.02**	**99.98±0.04**	**99.92±0.13**	**100.00±0.00**	**99.52**

3) Comparison with other deep learning models

To comprehensively measure the performance of the proposed method, we use other advanced deep learning models for comparative analysis, including SAE [12], DBN [3], Concept Drift: SEA [19], ACWGAN-GP [36] and ACGAN-SN [37]. ACWGAN-GP and ACGAN-SN are implemented in the proposed method by removing the DAM. The performance and training time of each model are shown in Fig. 5. It can be seen that the training time of ACWGAN-GP and ACGAN-SN is too long, which is one of the reasons why the GAN-based method is not used in this paper. The proposed method achieves the best results in three performance indicators, and has high computational efficiency. Overall, the proposed method has great advantages.

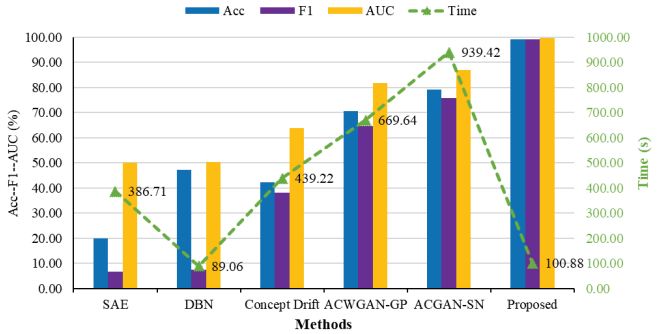


Fig. 5. Results for different methods on Canada data

E. Case 2 CWRU datasets

1) Experiment setup

The dataset comes from the CWRU. The sampling frequency at the driving end is 12 kHz. There are 10 state types and the data of three different working conditions are obtained at 1772 rpm, 1750 rpm and 1730 rpm, which are named dataset F, dataset G and dataset H, respectively.

The data imbalance setting method is similar to that of [20]. Among them, the first 50% of datasets F, G and H are divided into training sets in a certain proportion, which are recorded as F1, G1 and H1, and the last 50% of data are divided into test sets, which are recorded as F2, G2 and H2. The details are shown in Table V.

2) Results and analysis

We use the same method as Case1 to verify the classification performance of the proposed method. The test rule is cross-test uses different speeds, e.g., F1, G1→G2 denotes the transfer task from condition F to condition G, where F1 labels are known and G1, G2 labels are unknown.

Table. VI shows the testing results of the proposed and compared methods in various domain transfer tasks. It can be seen that the average diagnostic accuracy of TICNN and RFFDN is 71.50% and 71.44%, respectively. The state-of-the-art DNCNN diagnosis accuracy is 87.70%. It is worth noting that DNCNN achieves 95.26% for F1, G1→G2, 99.98% for G1, H1→H2, and less than 87.96% for other tasks. We believe the reason behind this may be that the DNCNN handles the low-speed (G, H) to high-speed (F) transfer task with negative migration. The RFFDN and TICNN models do not extract enough features, resulting in weak generalization ability. The average diagnostic accuracy, F1, AUC of the proposed method are 97.76%, 97.57%, and 98.75% in all tasks. Combined with the previous analysis, we can see that these findings are a powerful argument for determining the efficiency of the proposed method.

TABLE V
THE DETAILS OF EACH DATASET IN CASE 2

Fault location	NC	BF			IF			OF		
Labels	0	1	2	3	4	5	6	7	8	9
Fault Diameter (mm)	0	0.007	0.014	0.021	0.007	0.014	0.021	0.007	0.014	0.021
Training dataset F1	50%	30%	30%	30%	10%	10%	10%	5%	5%	5%
Training dataset G1	50%	30%	30%	30%	10%	10%	10%	5%	5%	5%
Training dataset H1	50%	30%	30%	30%	10%	10%	10%	5%	5%	5%
Testing dataset F2/G2/H2	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%

TABLE VI
CLASSIFICATION RESULTS ± STANDARD DEVIATIONS (%)BASED ON DIFFERENT FEATURES AND MODELS IN CASE 2

Cross test	Metrics	F1, G1→G2	F1, H1→H2	G1, F1→F2	G1, H1→H2	H1, F1→F2	H1, G1→G2	Average
Method 1	Acc	75.57±19.95	77.68±13.88	67.33±13.04	54.71±9.56	73.24±6.49	66.01±7.86	69.09
(SMOTE)	F1	71.19±23.11	72.65±18.35	61.50±14.14	46.82±11.34	67.60±8.94	60.74±9.38	63.41
	AUC	86.43±11.08	87.60±7.71	81.85±7.24	74.84±5.31	85.13±3.60	81.12±4.37	82.82
Method 2	Acc	93.71±5.12	87.12±13.89	84.91±14.96	81.76±16.10	85.58±14.53	82.55±18.20	85.93
(CB Loss)	F1	91.58±7.06	85.90±14.39	81.78±17.71	81.11±17.11	82.84±17.63	79.35±22.06	83.76
	AUC	96.51±2.84	92.84±7.71	91.62±8.31	89.87±8.94	91.99±8.07	90.30±10.11	92.18
Method 3	Acc	96.49±3.89	92.60±9.32	87.86±8.52	80.74±7.63	95.41±3.94	69.55±8.43	87.10
(Focal loss)	F1	95.81±5.30	92.20±9.13	84.12±10.99	76.56±7.83	94.52±5.31	66.66±11.20	84.97
	AUC	98.05±2.16	95.89±5.18	93.25±4.73	89.30±4.23	97.45±2.19	83.08±4.68	92.83
Method 4	Acc	82.48±11.17	72.11±21.69	69.36±22.85	61.11±9.33	49.19±17.41	60.73±7.63	65.83
(MLP)	F1	81.76±11.12	68.32±24.04	63.25±26.00	56.49±10.88	43.64±18.88	54.01±11.13	61.24
	AUC	90.26±6.20	84.51±12.05	82.98±12.69	78.39±5.18	71.77±9.67	78.18±4.24	81.01
Method 5	Acc	95.26±10.59	83.60±10.12	79.54±11.93	99.98±0.03	79.89±10.13	87.96±8.41	87.70
(DNCNN)	F1	94.23±12.88	80.50±10.28	77.31±13.19	99.98±0.03	77.23±10.24	85.29±9.64	85.75
	AUC	97.36±5.88	90.89±5.62	88.63±6.62	99.99±0.01	88.83±5.62	93.31±4.67	93.16
Method 6	Acc	84.20±16.43	47.95±13.81	74.23±17.90	67.48±18.24	80.23±8.33	74.58±4.46	71.44
(RFFDN)	F1	81.97±20.06	39.45±16.80	70.54±20.76	63.33±22.05	77.83±9.77	71.87±4.11	67.49
	AUC	91.22±9.13	71.08±7.67	85.68±9.94	81.93±10.13	89.01±4.63	85.87±2.48	84.13
Method 7	Acc	71.95±24.88	76.83±13.92	61.83±20.30	75.40±17.90	75.98±14.76	67.03±21.58	71.50
(TICNN)	F1	68.12±25.95	74.74±15.21	58.21±21.95	72.97±19.47	71.56±17.05	63.86±21.45	68.24
	AUC	84.42±13.82	87.12±7.73	78.79±11.27	86.33±9.94	86.65±8.20	81.68±11.99	84.16
Proposed	Acc	**100.00±0.00**	**95.45±6.81**	**98.66±1.15**	**93.33±11.79**	**99.18±8.57**	**99.95±2.42**	**97.76**
	F1	**100.00±0.00**	**94.32±8.51**	**98.64±1.16**	**93.33±11.79**	**99.18±10.29**	**99.95±2.42**	**97.57**
	AUC	**100.00±0.00**	**97.47±3.78**	**99.25±0.64**	**96.29±6.55**	**99.54±4.76**	**99.97±1.34**	**98.75**

This article has been accepted for publication in IEEE Transactions on Industrial Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3217541

3) Comparison with other deep learning models

Similar to Case 1, in this section we use the same five advanced deep learning models for comparative analysis. The statistical results of Acc, F1, and AUC are shown in Fig. 6. The SAE model has the worst performance. Although the GAN-based method has better diagnostic performance than SAE and DBN, it has the longest computation time. The performance of concept drift is mediocre. Considering the diagnostic performance and computational efficiency of all models, the proposed method has obvious advantages.

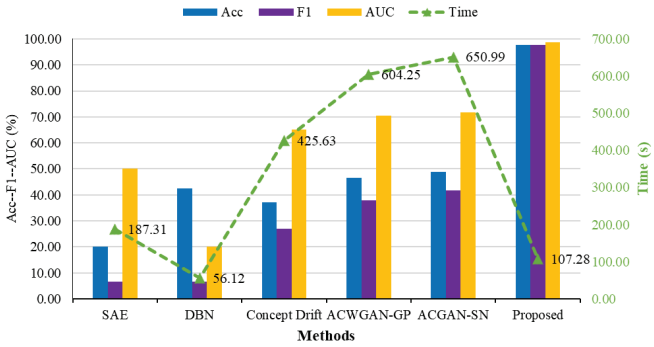


Fig. 6. Results for different methods on CWRU data

F. Analysis and Discussion

1) Parameter Analysis for SA

Ω1 and Ω2 are two key parameters of the proposed SA algorithm. Different parameter combinations will affect the subsequent diagnosis performance. Therefore, we tested the influence of these two hyperparameters on the performance of the model, and the results are shown in the Fig. 7. The results show that the values of Ω1 and Ω2 need a trade-off strategy; that is, when the values of Ω1 and Ω2 are larger, the distance between them cannot be too small. Ω1 and Ω2 determine the effect of the filter. When their values are close and their values are large, the filtered signal contains a large number of components independent of the fault characteristic frequency, resulting in a decline in the performance of the fault diagnosis model.

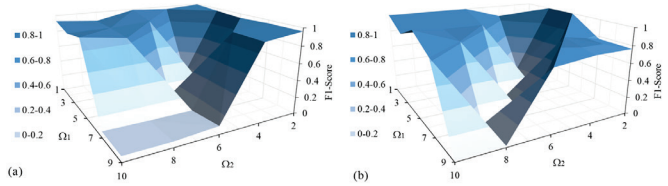


Fig. 7. Influences of Ω1 and Ω2 on (a) Case 1; (b) Case 2

2) Hyperparameter sensitivity analysis

Hyperparameter sensitivity experiments are discussed in this section. We use ACC, F1-score, and AUC to quantitatively evaluate the impact of Bs and Lr on the proposed model. The value range of Bs is [10,25,50,100,100,150,200,300], and the value range of Lr is [0.001,0.002,0.005,0.007,0.01,0.05,0.1] The experimental results of Case1 and Case2 are shown in Fig. 8. On the whole, when the two parameters are within a certain range, the proposed method can achieve satisfactory results. Based on experimental analysis, the hyperparameters of the two Cases are [Bs = 25, Lr = 0.002] and [Bs = 100, Lr = 0.007].

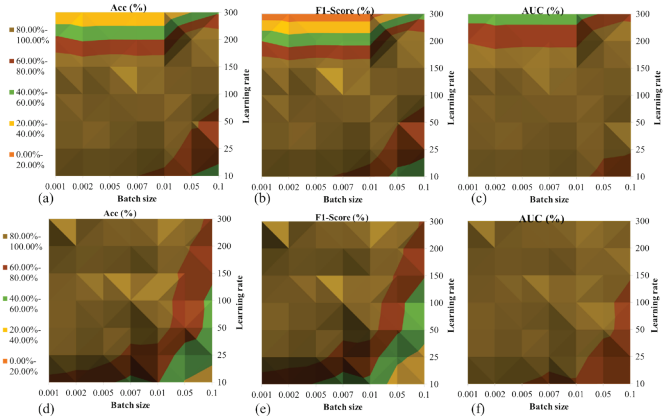


Fig. 8. Parameters sensitivity analysis. Experimental results of ACC, F1-Score, AUC for (a-c): Case 1 and (d-f): Case 2.

3) Complexity Analysis and Time Efficiency

In order to prove the simplicity of the proposed method, we refer to and use TC and SC (temporal and spatial complexity) for analysis. All experiments were completed in a laptop computer, with AMD Ryzen 5 4500u, 16G RAM, Python 3.8 We use Big-O notation [38] to describe the asymptotic upper bound of the order of magnitude of the function. The TC and SC calculation formulas of the convolutional and FC layers are as follows:

$$TC_{CNN} = O(D_{out} \cdot K \cdot C_{in} \cdot C_{out}) \tag{13}$$

$$SC_{CNN} = O(KC_{in} \cdot C_{out} + D_{out} \cdot C_{out}) \tag{14}$$

$$TC_{FC} = O(D_{out} \cdot w \cdot h \cdot D_{in}) \tag{15}$$

$$SC_{FC} = O(D_{out} \cdot D_{in}) \tag{16}$$

where, D_{out} is the characteristic dimension of the output node, D_{in} is the characteristic dimension of the input node, C_{in} and C_{out} are the number of input and output channels of convolutional layer, K is the kernel length. w and h denote the width and height of the input feature. Therefore, the TC and SC calculation formulas of the proposed method are:

$$TC_{Proposed} = \sum_{i=1}^{5} TC_{CNN} + \sum_{i=1}^{2} TC_{FC} \tag{17}$$

$$SC_{Proposed} = \sum_{i=1}^{5} SC_{CNN} + \sum_{i=1}^{2} SC_{FC} \tag{18}$$

According to the above formula, the TC and SC of the proposed method are 48900400 FLOPs and 35777 Bytes, respectively. The results of TC, SC, and time of each method are calculated by Big-O notation, as shown in Table VII. It can be seen that the proposed method is a relatively efficient model.

TABLE VII
COMPARISON OF ALGORITHM COMPLEXITY AND TIME EFFICIENCY

Methods	TC(FLOPs)	SC(Bytes)	Time(s)
MLP	94512500	3782405	102.90
DNCNN	309350400	70369	564.44
RFFDN	53287200	34961	128.54
TICNN	48171200	47385	104.79
SAE	81716800	3266757	386.71
DBN	33851550	1354757	89.06
Concept Drift	481129000	924770	439.22
ACWGAN-GP	24368339600	5797506	669.64
ACGAN-SN	24368339600	5797506	939.42
Proposed	48900400	35777	100.88

4) Statistical Test

We counted the diagnostic results of all models (13 models) involved in this paper on all datasets (all fault diagnosis tasks, a total of 14 datasets). The critical difference (CD) diagram was used for the analysis and the results are shown in Fig. 9. The thick horizontal lines indicate that the performance between adjacent models is comparable. The smaller the CD value, the higher the corresponding method performance. Obviously, the proposed method outperforms other comparative methods.

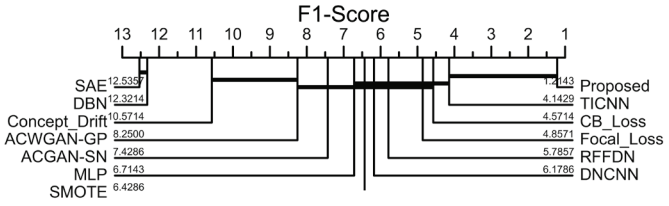


Fig. 9. CD diagram based on F1-Score.

5) Comparison with Related Works

To fully verify the high performance of the proposed method, we compare it with the related work under unbalanced and variable speeds. The comparison method includes the seven latest algorithms and a variant of the proposed method. The variants are OT+DAM+DTCNN, and OT (order tracking) [39] is the current popular speed elimination algorithm. Table VIII is a summary of the comparison results, in which accuracy is the worst accuracy obtained in the related literature.

The diagnostic accuracy of all methods was higher than 90%. Literatures [20, 34, 40-43] can only solve the problem under the condition of constant speed. Literatures [34, 41, 42] do not discuss fault diagnosis under domain shifts, which limits the application of speed shifts in practical engineering. Although the accuracy of reference [21] is higher than that of our algorithm, it is a fault diagnosis research for gearboxes, and it does not discuss the diagnosis effect of the model under constant speed conditions and domain shifts. The diagnostic accuracy of OT+DAM+ DTCNN is 71.81%, which indicates that the feature extraction ability of the OT algorithm is not as good as that of SA, and the resampling error is a possible reason.

On the whole, the proposed method solves the problems of constant speed, time-varying speed, domain shift and unbalanced samples in the same model for the first time and achieves a diagnosis accuracy of 97.76%. The superiority of the proposed method is fully proven.

6) Future research

Although the proposed method simulates the actual engineering application scenario as much as possible and achieves high diagnosis accuracy, there are still some problems to be solved in the future. We provide three possible research directions.

(1) Considering the feature extraction ability of SA and the ability to eliminate speed fluctuation. In the future, the SA algorithm can be applied to cross-domain fault diagnosis in different equipment, and the speed types of different equipment can be different.

(2) Given the influence of background noise, future work will add a noise reduction function on the basis of the proposed method to maintain the fault classification ability under strong noise.

(3) The ultimate goal of intelligent fault diagnosis is to integrate fault diagnosis knowledge into the model to realize the classification model without fault data.

TABLE VIII
COMPARISON WITH RELATED WORK

Algorithms	Datasets	Domain shift	Unbalanced data	Speed	Accuracy (the worst)
[41]	Two private datasets	$P=Q$	Yes	Constant	94.28%
[21]	Two gear datasets	$P=Q$	Yes	Time-varying	99.86%
[42]	CWRU; SQ dataset	$P=Q$	Yes	Constant	97.44%
[34]	UoC; Pu; JNU; SEU	$P=Q$	Yes	Constant	90.80%
[40]	CWRU; IMS	$P\neq Q$	Yes	Constant	95.83%
[20]	CWRU; Private data	$P\neq Q$	Yes	Constant	98.56%
[43]	CWRU; Private data	$P\neq Q$	No	Constant	96.18%
OT+DAM+ DTCNN	Canada	$P\neq Q$	Yes	Time-varying	71.81%
Proposed	**Canada; CWRU**	$P\neq Q$	**Yes**	**Time-varying; Constant**	**97.76%**

V. CONCLUSION

In this paper, a new method called SA-DTCNN is proposed for unbalanced fault classification of rolling bearings under various speeds conditions. The main conclusions are as follows.

First, the experimental results show that the SA algorithm has a strong feature extraction ability, and it can extract similar fault features from vibration signals under time-varying speed and constant speed. Then, the problem of unbalanced samples and various speeds can be solved by the cooperation of DAM. In addition, the performance of the proposed method is verified in two unbalanced datasets, including one time-varying speed and one constant speed datasets, which proves that SA-DTCNN has strong generalization ability.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, no. 7553, pp. 436-444, 2015/05/27 2015.

[2] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Sig. Process.,* vol. 115, pp. 213-237-237, 2019/01 2019.

[3] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Sig. Process.,* vol. 138, p. 106587, 2020/04/01/ 2020.

[4] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings," (in English), *Shock Vib.,* vol. 2017, 2017.

[5] G. Q. Zhao, K. K. Wu, Y. C. Gao, Y. M. Liu, and C. Hu, "Bearing fault diagnosis from raw vibration signals using multi-layer extreme learning machine," (in English), *Proceedings of 2019 14th Ieee International Conference on Electronic Measurement & Instruments (Icemi),* pp. 1287-1293, 2019.

[6] M. Saffari, M. Khodayar, S. M. J. Jalali, M. Shafie-khah, and J. P. S. Catalao, "Deep Convolutional Graph Rough Variational Auto-Encoder for Short-Term Photovoltaic Power Forecasting," presented at the 2021 International Conference on Smart Energy Systems and Technologies (SEST), 2021/09/06, 2021. [Online]. Available: http://dx.doi.org/10.1109/SEST50973.2021.9543326.

[7] D. Zhao *et al.*, "Enhanced data-driven fault diagnosis for machines with small and unbalanced data based on variational auto-encoder,"

Measurement Science and Technology, vol. 31, no. 3, p. 035004, 2019/12/30 2019.

[8] M. Khodayar, J. Wang, and M. Manthouri, "Interval Deep Generative Neural Network for Wind Speed Forecasting," *IEEE Trans. Smart Grid,* vol. 10, no. 4, pp. 3974-3989, 2019/07 2019.

[9] M. Khodayar, J. Wang, and Z. Wang, "Energy Disaggregation via Deep Temporal Dictionary Learning," *IEEE Trans. Neural Networks Learn. Syst.,* vol. 31, no. 5, pp. 1696-1709, 2020/05 2020.

[10] S. Zhao *et al.*, "A Review of Single-Source Deep Unsupervised Visual Domain Adaptation," *IEEE Trans. Neural Networks Learn. Syst.,* vol. 33, no. 2, pp. 473-493, 2022/02 2022.

[11] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Sig. Process.,* vol. 100, pp. 439-453, 2018/02 2018.

[12] L. Wen, L. Gao, and X. Li, "A New Deep Transfer Learning Based on Sparse Auto-Encoder for Fault Diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 49, no. 1, pp. 136-144, 2019/01 2019.

[13] J. Wu, T. Tang, M. Chen, Y. Wang, and K. Wang, "A study on adaptation lightweight architecture based deep learning models for bearing fault diagnosis under varying working conditions," *Expert Syst. Appl.,* vol. 160, p. 113710, 2020/12 2020.

[14] T. Zhang *et al.*, "Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions," *ISA Trans,* vol. 119, pp. 152-171, Jan 2022.

[15] F. Jia, Y. Lei, N. Lu, and S. Xing, "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization," *Mech. Syst. Sig. Process.,* vol. 110, pp. 349-367-367, 2018/09 2018.

[16] T. Pan, J. Chen, J. Xie, Z. Zhou, and S. He, "Deep Feature Generating Network: A New Method for Intelligent Fault Detection of Mechanical Systems Under Class Imbalance," *IEEE Trans. Ind. Inf.,* vol. 17, no. 9, pp. 6282-6293, 2021/09 2021.

[17] M. Khodayar, S. Mohammadi, M. E. Khodayar, J. Wang, and G. Liu, "Convolutional Graph Autoencoder: A Generative Deep Neural Network for Probabilistic Spatio-Temporal Solar Irradiance Forecasting," *IEEE Trans. Sustainable Energy,* vol. 11, no. 2, pp. 571-583, 2020/04 2020.

[18] J. C. Schlimmer and R. H. Granger, "Incremental Learning from Noisy Data," *MLear,* vol. 1, no. 3, pp. 317-354, 1986/09/01 1986.

[19] P. Ksieniewicz and P. Zyblewski, "Stream-learn — open-source Python library for difficult data stream batch analysis," *Neurocomputing,* vol. 478, pp. 11-21, 2022/03/14/ 2022.

[20] K. Xu, S. Li, X. Jiang, Z. An, J. Wang, and T. Yu, "A renewable fusion fault diagnosis network for the variable speed conditions under unbalanced samples," *Neurocomputing,* vol. 379, pp. 12-29, 2020.

[21] C. Wang, H. Li, K. Zhang, S. Hu, and B. Sun, "Intelligent fault diagnosis of planetary gearbox based on adaptive normalized CNN under complex variable working conditions and data imbalance," *Measurement,* vol. 180, 2021.

[22] Y. Cheng *et al.*, "An improved envelope spectrum via candidate fault frequency optimization-gram for bearing fault diagnosis," *J. Sound Vib.,* vol. 523, p. 116746, 2022/04/14/ 2022.

[23] D. Wei, T. Han, F. Chu, and M. J. Zuo, "Weighted domain adaptation networks for machinery fault diagnosis," *Mech. Syst. Sig. Process.,* vol. 158, 2021.

[24] H. Su, X. Yang, L. Xiang, A. Hu, and Y. Xu, "A novel method based on deep transfer unsupervised learning network for bearing fault diagnosis under variable working condition of unequal quantity," *Knowl-Based Syst,* vol. 242, 2022.

[25] X. N. Yu, Z. P. Feng, and M. Liang, "Analytical vibration signal model and signature analysis in resonance region for planetary gearbox fault diagnosis," (in English), *J. Sound Vib.,* vol. 498, Apr 28 2021.

[26] R. Zhen, Z. Shi, J. Liu, and Z. Shao, "A novel arena-based regional collision risk assessment method of multi-ship encounter situation in complex waters," *Ocean Engineering,* vol. 246, p. 110531, 2022/02 2022.

[27] A. S. Tarawneh, A. B. Hassanat, G. A. Altarawneh, and A. Almuhaimeed, "Stop Oversampling for Class Imbalance Learning: A Review," *IEEE Access,* vol. 10, pp. 47643-47660, 2022.

[28] M. A. Arefeen, S. T. Nimi, and M. S. Rahman, "Neural Network-Based Undersampling Techniques," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 52, no. 2, pp. 1111-1120, 2022/02 2022.

[29] Y. Qin, Q. Qian, J. Luo, and H. Pu, "Deep Joint Distribution Alignment: A Novel Enhanced-Domain Adaptation Mechanism for Fault Transfer Diagnosis," *IEEE Trans. Cybern.,* pp. 1-11, 2022.

[30] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep Transfer Learning with Joint Adaptation Networks," p. arXiv:1605.06636. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2016arXiv160506636L

[31] Z. Zhao *et al.*, "Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study," *ISA Trans.,* vol. 107, pp. 224-255-255, 2020/12 2020.

[32] H. Huang, N. Baddour, and M. Liang, "Multiple time-frequency curve extraction Matlab code and its application to automatic bearing fault diagnosis under time-varying speed conditions," *MethodsX,* vol. 6, pp. 1415-1432, 2019.

[33] H. Xue, M. Wu, Z. Zhang, and H. Wang, "Intelligent diagnosis of mechanical faults of in-wheel motor based on improved artificial hydrocarbon networks," *ISA Trans.,* vol. 120, pp. 360-371, 2022/01 2022.

[34] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Learning from Class-imbalanced Data with a Model-Agnostic Framework for Machine Intelligent Diagnosis," *Reliability Engineering & System Safety,* vol. 216, p. 107934, 2021/12/01/ 2021.

[35] J. K. Starling, C. Mastrangelo, and Y. Choe, "Improving Weibull distribution estimation for generalized Type I censored data using modified SMOTE," *Reliability Engineering & System Safety,* vol. 211, p. 107505, 2021/07 2021.

[36] Z. Li, T. Zheng, Y. Wang, Z. Cao, Z. Guo, and H. Fu, "A Novel Method for Imbalanced Fault Diagnosis of Rotating Machinery Based on Generative Adversarial Networks," *IEEE Trans. Instrum. Meas.,* vol. 70, pp. 1-17, 2021.

[37] Q. Tong *et al.*, "A Novel Method for Fault Diagnosis of Bearings with Small and Imbalanced Data Based on Generative Adversarial Networks," *Applied Sciences,* vol. 12, no. 14, p. 7346, 2022/07/21 2022.

[38] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5353-5360.

[39] G. Tang, Y. Huang, and Y. Wang, "Fractional frequency band entropy for bearing fault diagnosis under varying speed conditions," *Measurement,* vol. 171, p. 108777, 2021/02 2021.

[40] B. Zhao, X. Zhang, H. Li, and Z. Yang, "Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions," *Knowl-Based Syst,* vol. 199, p. 105971, 2020/07/08/ 2020.

[41] J. Miao, J. Wang, D. Zhang, and Q. Miao, "Improved Generative Adversarial Network for Rotating Component Fault Diagnosis in Scenarios With Extremely Limited Data," *IEEE Trans. Instrum. Meas.,* vol. 71, pp. 1-13, 2022.

[42] T. Zhang, J. Chen, F. Li, T. Pan, and S. He, "A Small Sample Focused Intelligent Fault Diagnosis Scheme of Machines via Multimodules Learning With Gradient Penalized Generative Adversarial Networks," *IEEE Trans. Ind. Electron.,* vol. 68, no. 10, pp. 10130-10141, 2021/10 2021.

[43] X. Wang, C. Shen, M. Xia, D. Wang, J. Zhu, and Z. Zhu, "Multi-scale deep intra-class transfer learning for bearing fault diagnosis," *Reliability Engineering & System Safety,* vol. 202, p. 107050, 2020/10/01/ 2020.

Feiyu Lu received the B.S. degree from Beihua University, in 2018, and the M.S. degree from Shijiazhuang Tiedao University, in 2021. He is currently pursuing the Ph.D. degree with Beijing Jiaotong University. His main research interests include rotating machinery condition monitoring and fault diagnosis.

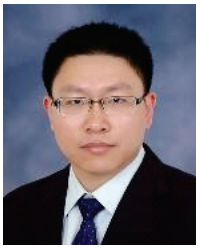
Qingbin Tong received the Ph.D. degree in Instrument Science and Technology from the Harbin Institue of Technology, Harbin, China, in 2008.

He is currently a Professor in the School of Electrical Engineering, Beijing Jiaotong University. He is mainly engaged in the research of artificial

intelligence and intelligent testing under rail transit, power and electronics, and energy, including the dynamic modeling of key components of the system; fault diagnosis, damage assessment and life prediction; dynamic nonlinear and non-stationary signal analysis and processing.

Ziwei Feng received the B.S. degree in Mechanical and electronic engineering from the Northeast Forestry University, Harbin, China, in 2021.

She is currently pursuing the Ph.D. degree with Beijing Jiaotong University. She is mainly engaged in the research of dynamic nonlinear and non-stationary signal analysis and processing.



Qingzhu Wan was born in Hubei, China, in 1975. He received M.Eng. in power system and its automation from the Harbin Institute of Technology, Harbin, China, in 2004, and the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2008. He is currently an associate professor with the school of Electrical and Control Engineering, North China University of Technology. His research interests include distributed generation, DC and AC microgrid, electricity markets and big data technology.