IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 16, 2018, accepted December 27, 2018, date of publication January 1, 2019, date of current version January 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2890693

Imbalanced Fault Diagnosis of Rolling Bearing Based on Generative Adversarial Network: A Comparative Study

WENTAO MAO[1,2], **YAMIN LIU**[1], **LING DING**[1], **AND YUAN LI**[1]

[1]School of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China
[2]School of Mechanics and Civil and Architecture, Northwestern Polytechnical University, Xi'an 710129, China

Corresponding author: Wentao Mao (maowt@htu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U1704158, Grant U1604154, and Grant 11702087, in part by the China Postdoctoral Science Foundation Special Support under Grant 2016T90944, and in part by the Foundation of Henan Normal University for Excellent Young Teachers under Grant 14YQ007.

ABSTRACT Due to the real working conditions and data acquisition equipment, the collected working data of bearings are actually limited. Meanwhile, as the rolling bearing works in the normal state at most times, it is easy to raise the imbalance problem of fault types which restricts the diagnosis accuracy and stability. To solve these problems, we present an imbalanced fault diagnosis method based on the generative adversarial network (GAN) and provide a comparative study in detail. The key idea is utilizing GAN, a kind of deep learning technique, to generate synthetic samples for minority fault class and then improve the generalization ability of the fault diagnosis model. First, this method applies fast Fourier transform to pre-process the original vibration signal and then obtains the frequency spectrum of fault samples. Second, it uses the spectrum data as the input of GAN to generate the synthetic minority samples following the data distribution of the real samples. Finally, it puts the synthetic samples into the training set and builds a stacked denoising auto encoder model for fault diagnosis. To testify the effectiveness of the proposed method, a series of comparative experiments is carried out on the CWRU bearing dataset. The results show that the proposed method can provide a better solution for imbalanced fault diagnosis on the basis of generating similar fault samples. As a comparative study, the proposed method is compared to several diagnostic methods with traditional time-frequency domain characteristics. Moreover, we also demonstrate that the proposed method outperforms three widely used sample synthesis techniques, such as random oversampling, synthetic minority oversampling technique, and the principal curve-based oversampling method in terms of diagnosis accuracy and numerical stability.

INDEX TERMS Generative adversarial network, fault diagnosis, imbalanced fault, SDAE.

I. INTRODUCTION

Generally speaking, rolling bearing is one of the most important components of mechanical equipment. Due to the complex structure and operating conditions, bearing tends to be damaged easily, causing huge damage even casualties. Therefore, it is a crucial task to monitor and diagnose rolling bearings fault. In most time of practical applications, rolling bearing works in normal state and relatively few fault data could be collected. Consequently, the data imbalance problem arises [1]. The so-called *data imbalance* refers to that the amount of fault data collected is far less than that of normal samples, which leads to the model bias of data-driven fault diagnosis methods. For example, if 90 normal samples and 10 fault samples are available, we can get the prediction accuracy of 90%, with all normal samples classified correctly while all fault samples being classified to normal class as well. This accuracy is relatively high but meaningless, which is also called "undo" performance. Therefore, it's of great significance to improve the prediction performance for data imbalance problem, especially on the minority class data.

Many research works have been put forward to solve the data imbalanced problem in bearing fault diagnosis from two aspects: data and algorithm. From the data perspective, Chawla *et al.* [2] proposed a Synthetic Minority Oversampling Technique (SMOTE) to randomly insert virtual samples for balancing the training set. In this kind of work,

the quality of the synthetic samples may not be good enough to provide useful information for imbalanced fault diagnosis. Thus, Ramentol *et al.* [3] put forward an improved SMOTE algorithm based on rough set theory to oversampling the minority samples. Gao *et al.* [4] combined SMOTE with particle swarm optimization (PSO) and RBF classifier for enhancing the importance of the minority samples. Mao *et al.* [5] proposed a principal curve-based method for oversampling minority samples and under-sampling majority samples respectively. From the algorithm perspective, the current works are devoted to improve the traditional algorithm structure or design new algorithms by means of data characteristics. Jia *et al.* [6] proposed a deep normalized convolutional neural network and a neuron activation maximization algorithm for imbalanced fault classification. Aiming at data imbalanced problem, Yin *et al.* [7] proposed a kernel fisher linear discriminant analysis approach. Sun *et al.* [8] added a cost item into AdaBoost learning framework to adjust the weight of the minority samples. Using similar idea, Xiong *et al.* [9] improved AdaBoost framework for the minority class analysis by using a local clustering ensemble learning method.

Although these methods mentioned above have obtained good performance, they still have some drawbacks, one of which is lack of adaptability, i.e., they are unable to learn the data distribution characteristic of samples automatically. Thanks to the quick development of deep learning techniques, Generative Adversarial Network (GAN), proposed by Goodfellow *et al.* [10] in 2014 firstly, has received wide attention in the fields of computer vision [11] and text analysis [12], etc. GAN is capable of learning the data distribution characteristic of the original samples and then generating new synthetic samples with similar distribution. Briefly speaking, GAN is composed of a generator and a discriminator both of which could be a normal fully-connected neural network. The generator mainly captures the data distribution, while the discriminator is used to estimate the probability that a sample comes from the training data rather than the generator. The training procedure is a process of game, in which the generative and discriminative ability of GAN will continue to rise up until the Nash equilibrium is reached. With the continuous improvement of GAN, its performance has been further promoted and got successful applications. For examples, Radford *et al.* [13] combined Deep Convolution Neural Network (DCNN) with GAN to improve the stability of the training process and the quality of the generation. Arjovsky *et al.* [14] introduced Wasserstein distance to solve the drawbacks like difficult convergence and mode collapse in GAN's training process. As a result, the trained GAN can generate more similar synthetic images to enlarge the amount of minority samples. However, according to our literature research, GAN is still in its infancy in the field of fault diagnosis. Among these works, Wang *et al.* [15] combined GAN with Stacked Denoising Auto Encoder (SDAE) to perform gearbox fault diagnosis. GAN was used to expand the sample amount, and SDAE performed as a discriminator of GAN to extract deep features adaptively and diagnose fault types as well. Focusing on the bearing fault diagnosis, Lee *et al.* [16] utilized the synthetic samples generated through GAN from the Empirical Mode Decomposition (EMD) energy spectrum data, and then obtained better fault diagnosis results than the traditional over-sampling technologies. Liu *et al.* [17] proposed a deep neural network based on GAN to conduct fault diagnosis, which trains an Auto Encoder through an adversarial training process and imposes a prior distribution on the latent coding space. Han *et al.* [18] introduced adversarial learning as a regularization into Convolution Neural Network (CNN) which makes the feature representation robust and boosts the generalization ability of the trained model.

Based on the above analysis, it is clear the key issue to improve the effect of imbalanced fault diagnosis is how to exploit the data distribution of minority fault samples and generate good synthetic samples adaptively. Although the aforementioned works all get satisfactory performance in fault diagnosis, they still have not gave detailed instruction of applying GAN in the aspects of selection of data source and feature extraction method as well as the effect of oversampling, etc. Following this idea, this paper tries to provide a comparative study and give a detailed guidance of applying GAN in imbalanced fault diagnosis. Specifically, this paper presents a method which uses GAN to expand the capacity of the minority samples adaptively and then adopts SDAE to extract deep feature and conduct final diagnosis. We must point out that, as this method is merely a direct combination of two new machine learning algorithms, some variants of GAN and SDAE also can be introduced and we will not mention it specifically. We design two kinds of experiments on the CWRU bearing dataset, not only verifying the quality of GAN synthetic samples, but also comparing the classification performance with some state-of-the-art methods. And the comparative results demonstrate the effectiveness of GAN in solving the imbalanced fault diagnosis problem.

The remainder of the paper is organized as follows. We introduce the basic theory of GAN and SDAE in Background section. The third part is the method proposed in this paper. The fourth section shows the simulation experiment and the results on CWRU dataset. Finally, we analyze and summarize the whole paper in the last part.

II. BACKGROUND

As a comparative study, we first give some brief introduction of the two main algorithms, SDAE and GAN.

A. SDAE

We first introduce Denoising Auto Encoder (DAE), and SDAE can be viewed as multiple DAEs stacked together. DAE which was proposed by Vincent *et al.* [19] in 2008 is a more robust auto-encoder for feature extraction. Like classical auto encoder, DAE consists of input layer, hidden layer and output layer. But different from classical auto encoder, DAE tries to enhance the robustness of the extracted feature

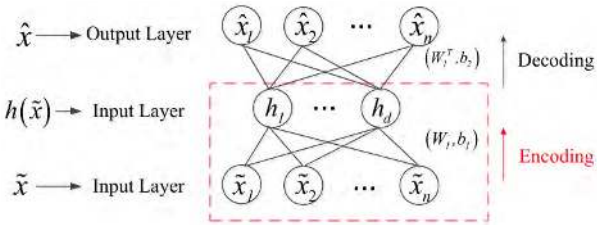

FIGURE 1. Structure of DAE.

by adding a little noise in the input layer. The structure of DAE is shown in Fig. 1.

The input of DAE is $\tilde{x}$, with noise added, and the output of the hidden layer can be expressed as:

$$h(\tilde{x}) = \sigma(W_1\tilde{x} + b_1) \tag{1}$$

where W_1 is the weight matrix from input layer to hidden layer, b_1 is the bias, and the activation function $\sigma(\cdot)$ is generally chosen as Sigmoid function. Obviously, the function of W_1 and b_1 can be summarized as encode of raw data. And DAE decodes the output of hidden layer by using the following equation:

$$h\left(\hat{x} = W_1^T h(\tilde{x}) + b_2\right) \tag{2}$$

Like the traditional Auto Encoder, the aim of DAE is to reproduce the input signals with as small bias as possible. Gradient descent algorithm is adopted to minimize the reconstruction error of DAE. The objective function can be represented as:

$$h\left(\tilde{x}_i, \hat{x}_i\right) = \frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{x}_i - \hat{x}_i\right\|^2 \tag{3}$$

By stacking multiple DAEs, SDAE can provide deep and robust feature extraction from raw input data, as shown in Fig. 2. The output of previous DAE performs as the input of next DAE, and the whole SDAE can be trained in the same manner.

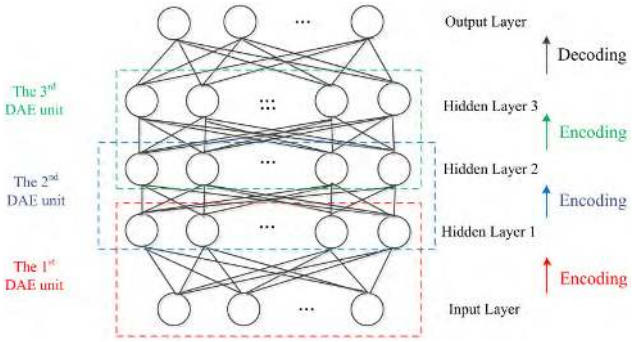

FIGURE 2. Structure of SDAE.

B. GAN

Inspired by a zero-sum game, GAN is composed of a generator and a discriminator, denoted by G and D respectively. The purpose of G is to generate synthetic samples whose data distribution are more similar to the real ones, while the aim of D is to discriminate the samples as True or False. For the sake of understanding, G is often compared to a counterfeit bill maker who aims to make as realistic a counterfeit bill as possible. On the contrary, D is often viewed as a policeman, who tries his best to recognize the bill's true or false. How to reach each aim is a process of game. Please note that G and D both can be any generative and discriminative algorithm respectively, including neural network. Fig. 3 gives a sketch map of GAN.

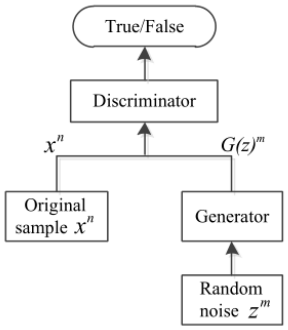

FIGURE 3. Sketch map of GAN.

In the specific training process, the goal of G is to learn the data distribution of real samples and generate synthetic samples that can't be distinguished from the real ones as far as possible. Specifically, the input of G is a set of random noise $Z = (z^1, z^2, \cdots, z^m)$. The output of G is the synthetic samples $G(Z) = (G(z)^1, G(z)^2, \cdots, G(z)^m)$, which has the similar distribution with the real one. The input of D is the synthetic samples $G(Z)$ or the real samples $X = (x^1, x^2, \cdots, x^n)$, and the goal of D is to identify the true and false samples by outputting the true or false logical value. During the training process, G and D are trained alternately until reach the Nash equilibrium, the loss function is shown in Equation (4) and (5).

$$\begin{aligned} L(D, G) = {} & E_{x\sim P_{data}(x)}\left[\log D(x)\right] \\ & + E_{z\sim P_z(z)}\left[\log\left(1 - D(G(z))\right)\right] \end{aligned} \tag{4}$$

$$\min_G \max_D L(D, G) \tag{5}$$

III. IMBALANCED FAULT DIAGNOSIS BASED ON GAN AND SDAE

To solve the imbalanced fault diagnosis problem, this paper presents a new diagnosis method by combining GAN and SDAE. Getting help from GAN, the method learns the data distribution adaptively and then generates similar samples to balance the majority and the minority classes. With the synthetic samples, SDAE is applied to feature extraction and final diagnosis. Fig. 4 shows the flow chart of this method.

A. DATA ACQUISITION AND PRE-PROCESSING

In the process of feature extraction for bearings, considering the complexity of original vibration signal, a common method [20] is to carry out Fast Fourier Transform (FFT) for

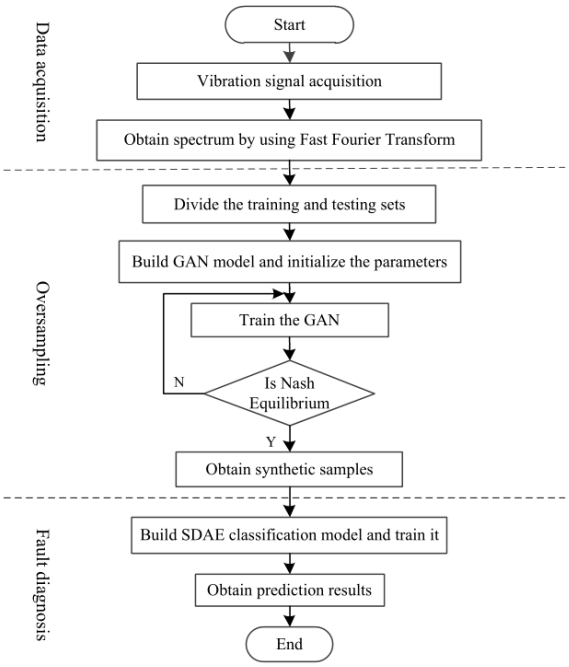

FIGURE 4. Flow chart of the proposed method.

the vibration signal, obtain signal's frequency domain representation, and then calculate the statistical characteristics to build heterogeneous feature. Compared to the extracted feature, the frequency spectrum contains more useful information about original signal. For instance, frequency spectrum is able to analyze composition of original signal quantitatively. Conversely, some prior information have been dropped out in statistical calculation. Therefore, in this paper frequency spectrum data are selected as the objects of GAN to generate synthetic sample. First, the original vibration signals in time domain are acquired from the test stand. Then we get the frequency spectrum samples $\{x_i, y_i\}_{i=1}^{N}$ by FFT, where N is the number of samples. Next, we divided the samples into training set and test set.

B. OVERSAMPLING OF FAULT SAMPLE USING GAN

Focusing on the imbalanced fault diagnosis problem, we apply GAN to oversampling the minority fault samples. Specifically, we use the frequency spectrum of minority fault samples as real sample in GAN, and then generate the synthetic spectrum samples whose distribution is similar to the original samples. Here, the G and D in GAN are both three-layer fully-connected neural networks. The neuron nodes of the hidden layer are 128 and 256 respectively, and the activation function is sigmoid function. The schematic is shown in Fig. 5.

The input of G is a random noise $Z = \left(z^1, z^2, \cdots, z^m\right)$ with the P_g data distribution. Its output is synthetic samples $G(z)$, whose data distribution is similar to the real samples P_{data}. The input of D is synthetic samples $G(z)$ or real samples X, and the output is the probability that the input of D is real sample. The two networks reached Nash equilibrium when

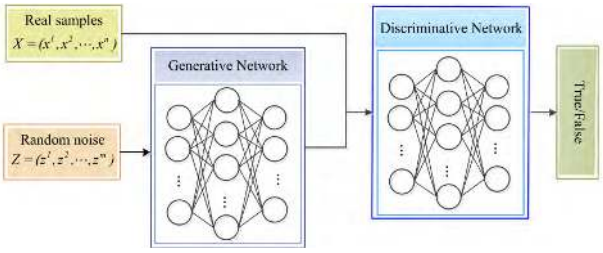

FIGURE 5. Schematic of GAN used in this paper.

the output is 0.5. G and D are trained alternately and they are trained as the follow steps:

Step 1: Initialize G and fixed it, then train D by maximizing the loss function in Equation (6). The discriminator is upgraded by using stochastic gradient ascending.

$$\max_D L(D) = \log D\left(x^{(i)}\right) \tag{6}$$

Step 2: Fix the D trained in Step 1, then upgrade G by minimizing its loss function in Equation (7) via descending the stochastic gradient.

$$\min_G L(G) = \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \tag{7}$$

Step 3: Determine whether the two networks have reached Nash equilibrium. If they do, the training process ends. Otherwise, repeat step 1 and step 2 alternately until they do. The detailed training process of GAN are shown in the Table 1 [10].

TABLE 1. The detailed training process of GAN.

Algorithm GAN Minibatch stochastic gradient descent training of generative adversarial nets. The number of steps to the discriminator, k, is a hyper parameter. We used k=1, the least expensive option, in our experiments.

for number of training iterations do
  for k steps do
    Sample minibatch of m noise samples $\{z^1, z^2, \cdots, z^m\}$ from noise prior $P_z(z)$.
    Sample minibatch of m examples $\{x^1, x^2, \cdots, x^m\}$ from data generating distribution $P_{data}(x)$.
    Upgrade the discriminator by ascending its stochastic gradient:
    $\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[logD\left(x^{(i)}\right) + log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right)\right]$.
  end for
  Sample minibatch of m noise samples $\{z^1, z^2, \cdots, z^m\}$ from noise $P_z(z)$.
  Upgrade the generator by descending its stochastic gradient:
  $\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right)\right]$
  end for
The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.

C. BUILD FAULT DIAGNOSIS MODEL USING SDAE

After training, GAN can generate the synthetic spectrum samples $G(z)$ whose distribution is similar to the real minority fault samples. As a result, the original samples $\{x_i, y_i\}_{i=1}^{N}$. Here $\tilde{x}_i$ is i-th sample of the balanced sample set $\tilde{X} = (x^1, x^2, \cdots, x^n, G(z)^1, G(z)^2, \cdots, G(z)^m)$. Then we feed $\tilde{X}$ into SDAE by setting it as the input of first DAE unit. From Equation (1), we can get the output $h_1 = \sigma(W_1\tilde{x} + b_1)$ of the

first DAE unit and then set it as the input of the next unit. Repeat this process until reach the last Softmax layer.

In this paper we directly use Softmax function to conduct classification for normal class and fault class. We also can extract the final layer as feature, and then introduce an independent classifier to conduct fault diagnosis. Here the classifier model can be Support Vector Machine (SVM), Extreme Learning Machine (ELM), Random Forest (RF) and any other classification algorithms. Due to space limitation, here we won't provide detailed analysis for these algorithms.

IV. EXPERIMENTS

To verify the effectiveness of the proposed method, two experiments are designed in this section. The Experiment 1 is a verification experiment under the condition of data balanced. The target here is to check the quality of two kinds of GAN synthetic samples, i.e. generated from frequency spectrum sample and heterogeneous feature sample. For convenience, we call these two kinds of synthetic samples as *spectrum synthetic samples* and *feature synthetic samples*, respectively. In Experiment 1, we run Test A and Test B to evaluate the quality of these two kinds of generated samples, and check which one is more suitable for fault diagnosis. On the base of the results of Experiment 1, we run the Experiment 2 to testify the performance of the proposed method. Experiment setting and the corresponding targets are listed in Table 2.

TABLE 2. Settings of two experiments and their goals.

Experiment		Target
1	Test A	Evaluate the quality of spectrum synthetic samples generated by GAN
	Test B	Evaluate the quality of feature synthetic samples generated by GAN
2		Utilize the synthetic samples generated by GAN for imbalanced fault diagnosis

A. CWRU DATASET INTRODUCE

CWRU bearing fault dataset [21] are obtained from the electrical engineering laboratory of Case Western Reserve University, USA. The electric spark is used to create damage on the inner race, outer race and ball of the bearing with different damage diameter such as 0.007, 0.014, 0.021 and 0.028 inches. Then the various vibration signals are collected under the different loads from 0 to 3 hp. Besides normal condition, this dataset adopts the fault data with sampling frequency of 12 kHz at fan end (FE) and drive end (DE) as well as the fault data with sampling rates of 48 kHz at DE. Therefore, this dataset contains four kinds of health condition: normal condition, inner race fault, outer race fault and ball fault.

B. EXPERIMENT 1

To evaluate the quality of two kinds of synthetic samples, Experiment 1 adopts 7 kinds models for comparison: ELM [22], Sparse Bayesian ELM (SBELM) [23], RF[24], Output Kernel Learning (OKL) [25], SVM [26], Deep Belief Network (DBN) [27] and SDAE [28]. Among of them, ELM, SBELM, RF, OKL and SVM are shallow models, while DBN and SDAE are deep models. For the synthetic samples generated from frequency spectrum, the input of ELM, SBELM, RF, OKL and SVM are all 59-dimensional features [29] in time and frequency domain, while DBN and SDAE extract deep features directly from frequency spectrum including original spectrum samples plus synthetic samples. For the synthetic samples generated from heterogeneous features, we only compare the diagnosis performance on five shallow models, with no using deep models for comparison.

The 59-dimensional features [29] used here are obtained from bispectrum analysis, time and frequency domain analysis, EMD, Wavelet Packet Decomposition (WPD) and envelope analysis. The extraction methods and the dimension of corresponding features are listed in Table 3.

TABLE 3. Extraction method and their feature dimension.

Feature extraction	Feature dimension
Bispectrum analysis	10
EMD	10
time and frequency domain analysis	13
WPD	8
Envelope analysis	18

1) SETTING AND ANALYSIS OF EXPERIMENT 1

In Experiment 1, we take ball fault and outer race fault as example. With the sampling frequency of 48 kHz and the load of 1 hp, 200 samples with damage radius of 0.007 inch, 0.014 inch and 0.021 inch are selected, respectively. Then we have 3 kinds of ball fault and 3 kinds of outer race fault, as shown in Table 4. And we visualize their time-domain vibration signals of selected samples in Fig. 6. For each subfigure, we randomly choose 2048 points.

TABLE 4. Information of samples used in Experiment 1.

Fault Type	Load (hp)	Damage radius (inches)	Number
Ball(Ball-1)	1	0.007	200
Ball(Ball-2)	1	0.014	200
Ball(Ball-3)	1	0.021	200
Outer race(Outer-1)	1	0.007	200
Outer race(Outer-2)	1	0.014	200
Outer race(Outer-3)	1	0.021	200

Fig. 7 illustrates the spectrum synthetic samples generated by GAN in Test A. For better comparison, we also show the real frequency spectrum samples from which GAN learns.

In Figure 7, it is clear that the synthetic samples are totally similar with real spectrum samples. In spite of small deviation at some peaks, we observe that the synthetic samples can follow the overall trend of real samples. In fact, we don't think it's preferable to make synthetic samples exactly identical to the real ones. Just like little noise for SDAE, such small difference between synthetic and real ones would improve

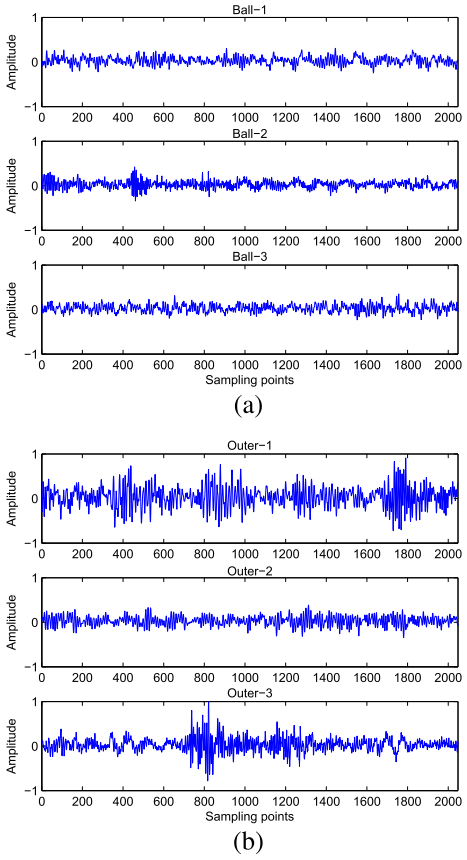

FIGURE 6. Time domain charts of vibration signals with (a) three ball fault types and (b) three outer fault types.

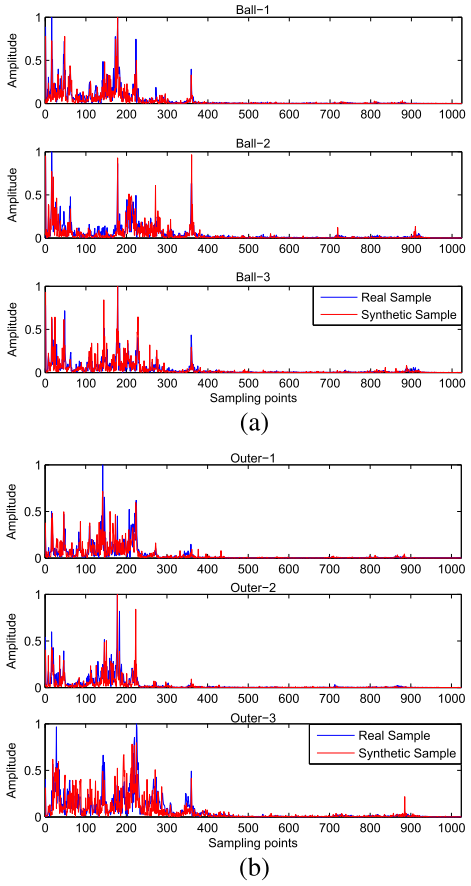

FIGURE 7. Comparison of the real spectrum samples and the synthetic samples generated by GAN with (a) three ball fault types and (b) three outer fault types.

the robustness of fault diagnosis model and enhance the generalization performance as well, which will be verified by the following comparative results.

To further check the influence of synthetic samples, Fig. 8 visualizes the data distribution of 59-dimensional features with different number of synthetic samples added in the training set. The X-axis and Y-axis are the main two-dimensional components obtained by TSNE [30], respectively.

From Fig. 8, with new synthetic samples added, the data distribution of different fault types is further separated, especially in Fig. 8 (c) and (d). Although more samples are introduced, the clustering effect of features is better obvious. Please compare Fig. 8 (c) and (d) in terms of blue and red points. The blue points in Fig. 8 (d) isolate from other color points apparently. Therefore, it's indicated that synthetic samples bring more discriminative information for classification.

For Test B, we use GAN to generate feature synthetic samples directly from the 59-dimensional features listed in Table 3. The comparison of real feature samples and the synthetic samples generated by GAN are shown in Fig.9.

In Fig. 9 (a) and 9(c), the synthetic samples by GAN seem to be similar to the real samples besides at the peak. However, from the local enlarged comparison in Fig. 9 (b) and 9 (d), it is clear that the synthetic samples, quite unlike the real samples, almost tend to be zero. That means GAN is unable to learn the characteristic information of the feature samples. We notice that the 59-dimensional features are discrete and rather short, and more importantly, of no relatedness between them, which doesn't like the frequency spectrum data. From GAN theory [10], it's hard to exploit the specific prior information about these fault samples, which implies the reason that the feature synthetic samples are quite divergent from the real samples. Comprehensively, the synthetic samples generated from frequency spectrum look better than the ones from 59-dimensional feature in terms of data shape.

We further verify the diagnosis performance using these synthetic samples in Test A & B. We firstly divide the original data set into training set and test set in proportion of 5:5, and then add in training set 300 synthetic samples for each fault type. The comparative results about test accuracy using seven classification models are shown in Fig. 10. Here SVM, SBELM and OKL all use 5-fold cross validation to find optimal parameters. The SDAE has two hidden layers with 50 and 30 neuron nodes respectively. The DBN has three hidden layers, and the neuron nodes are 512, 128, and 64. The number of ELM's hidden neuron is set 500 with hardlim activation function. The kernel function of SVM is RBF kernel while the parameter is set by cross validation.

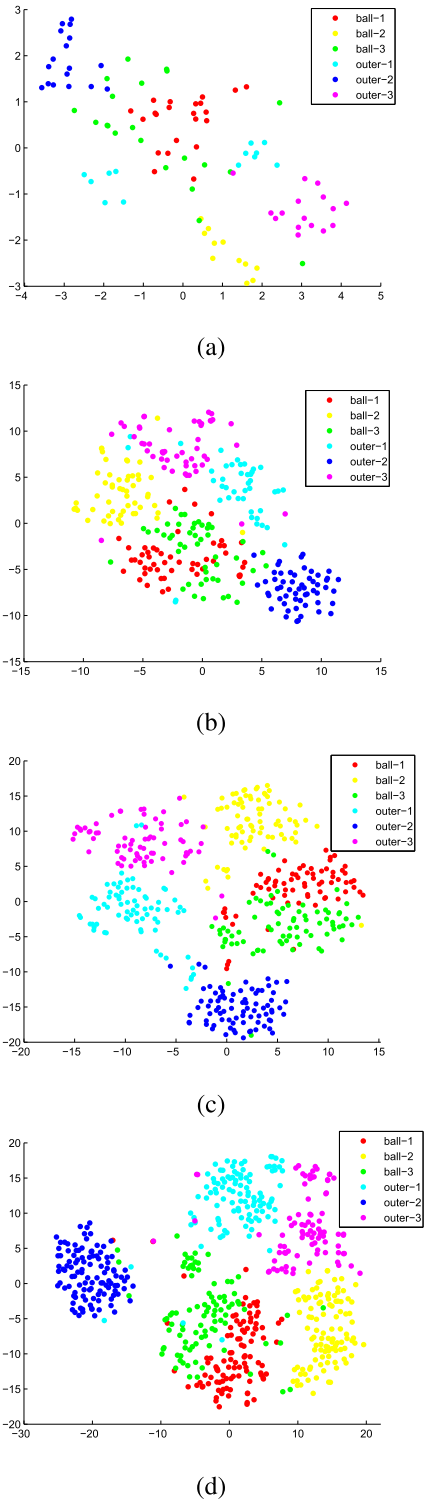

FIGURE 8. Visualized feature distribution using TSNE technique with different number of synthetic samples added, including (a) no synthetic samples, (b) each fault type added 200 synthetic samples, (c) each fault type added 350 synthetic samples, (d) each fault type added 500 synthetic samples. For better display effect, we here randomly choose one-sixth samples for illustration.

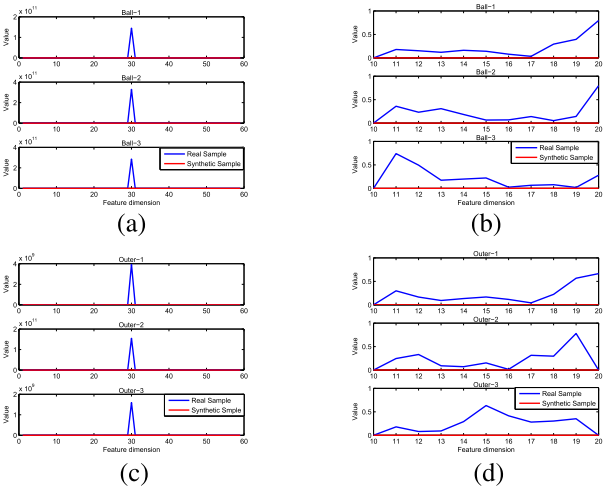

FIGURE 9. Comparison of the real feature sample and feature synthetic sample generated by GAN. Here (a) and (c) are the total comparative effect of ball fault and outer race fault respectively. (b) and (d) the local enlarged effect of (a) and (c) between the dimension 10-20 respectively.

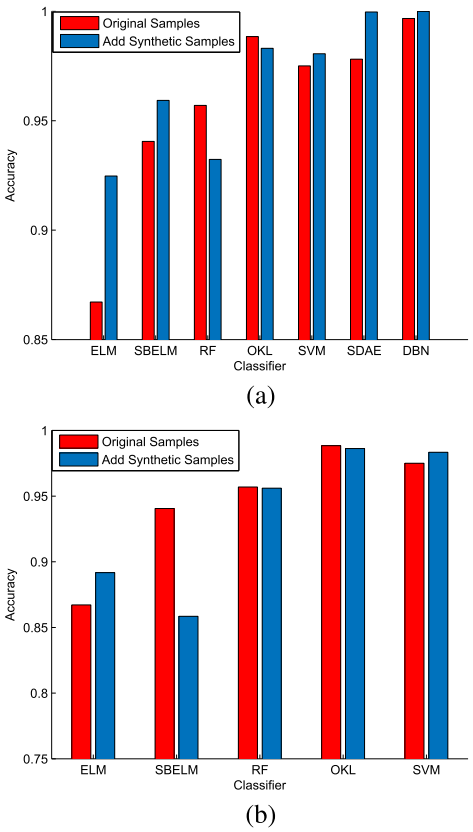

FIGURE 10. Test accuracy of different classifiers before and after adding 300 synthetic samples in (a) Test A and (b) Test B.

From Fig. 10(a), we find that the test accuracy of ELM, SBELM, SVM, SDAE and DBN are indeed risen up after 300 synthetic samples having been added for each fault type. It indicates that the synthetic spectrum samples by GAN could provide useful information for model training to some extent. However, for RF and OKL, the accuracy declined. With a superficial analysis, we think the reasons may be: 1) the feature used in this experiment is redundant for RF and

TABLE 5. Comparative results of Test A with different number of synthetic samples added.

Add Number	0	100	200	300	400	500
OKL	0.9884	0.9868	0.9842	0.9831	0.9842	0.9823
SVM	0.9751	0.9819	0.9806	0.9806	0.9809	0.9776
RF	0.957	0.9401	0.9368	0.9323	0.927	0.9294
SBELM	0.9406	0.947	0.9533	0.9593	0.9631	0.9667
ELM	0.8672	0.9221	0.9249	0.9247	0.9232	0.9215
SDAE	0.9781	0.9997	0.9998	0.9997	1	1
DBN	0.9967	1	1	1	1	1

TABLE 6. Comparative results of Test B with different number of synthetic samples added.

Add Number	0	100	200	300	400	500
OKL	0.9884	0.9897	0.9878	0.9862	0.9838	0.9794
SVM	0.9751	0.9821	0.9822	0.9834	0.9847	0.9852
RF	0.957	0.9582	0.9577	0.9561	0.9553	0.9577
SBELM	0.9406	0.8365	0.8423	0.8585	0.8744	0.8849
ELM	0.8672	0.8875	0.8753	0.8917	0.8883	0.8889

OKL; 2) the quality of generated spectrum samples isn't good enough to provide auxiliary information.

From Fig. 10(b), the accuracy of SBELM, RF and OKL all reduce after adding the feature synthetic samples. Especially for SBELM, the accuracy falls heavily. This reduction keeps line with Fig. 9 in which the feature synthetic samples by GAN look distinctly different from the real samples. This comparison further demonstrates that generating synthetic samples by GAN from 59-dimensional feature is not very helpful for fault diagnosis.

Also from Fig. 10(a), we observe an interesting phenomenon. The accuracies of two deep learning algorithms, i.e., SDAE and DBN, both have been significantly improved after adding synthetic samples. That's because the deep learning algorithm can extract adaptively rich and representative features from raw data, avoiding the disadvantages about insufficient or redundant handcrafted features.

We further add different number of synthetic samples into the training data set, and provide the comparative results of all seven classifiers, as in Table 5 and 6. In order to ensure the confidence of the results, the sample expansion is only carried out in the training set, and the test set is composed of the real samples all the time. To make a stable comparison, we take the mean value of 30 trials as the final result.

For better comparison, we plot the results in Table 5 and 6 in Fig. 11. Here the X-axis is the amount of added synthetic samples, and the Y-axis is test accuracy.

From Table 5 and 6 plus Fig. 11, we find that for spectrum synthetic samples (Test A), the test accuracy of ELM, SBELM, DBN and SDAE all rise up. It demonstrates that the spectrum synthetic samples have provided auxiliary and useful information for model training. Meanwhile, compared with these models, the accuracy of SVM is quite stable. However, the accuracy of RF and OKL both tend to decline along with the number of synthetic samples increasing. And for feature synthetic samples (Test B), only SVM and ELM

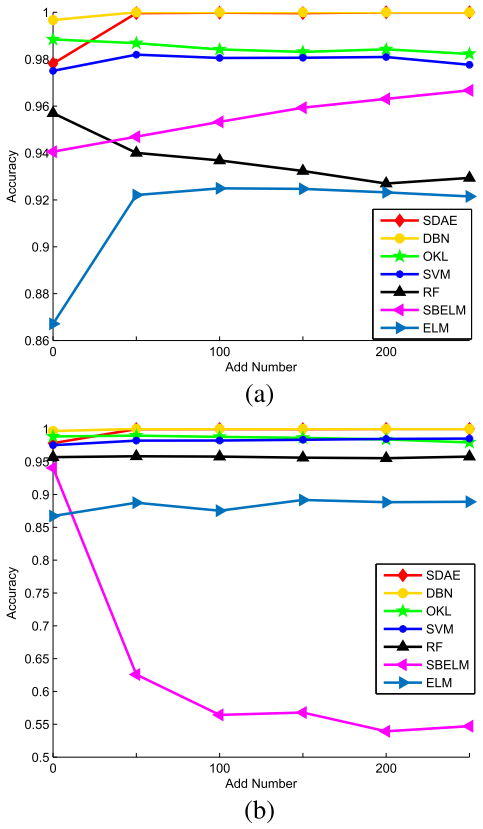

FIGURE 11. Test accuracy with different number of synthetic samples added in (a) Test A and (b) Test B.

have ascending lines. Same to Fig. 10, the accuracy of SBELM descends dramatically. These comparative results prove again that the frequency spectrum data can serve better than 59-dimensional feature as the source data learnt by GAN to generate synthetic samples for fault diagnosis.

TABLE 7. Standard deviation of 30 trials with different number of synthetic samples added in Test A.

Standard deviation	0	100	200	300	400	500
OKL	0.004	0.0048	0.0046	0.0044	0.0039	0.0047
SVM	0.0059	0.0021	0.0035	0.0056	0.0023	0.0066
RF	0.0091	0.0085	0.0106	0.0105	0.0106	0.0117
SBELM	0.0094	0.0088	0.0066	0.0053	0.0043	0.0039
ELM	0.019	0.0106	0.0092	0.0115	0.0093	0.0094

We also verify the numerical stability in Test A and B. Table 7 and 8 respectively give the standard deviation of 30 repeated trials with different number of synthetic samples added. Here to avoid the randomness of deep network, we only choose five shallow models for comparison.

Frankly speaking, the comparison effects in Table 7 and 8 are quite different. In Table 7, the overall standard deviation after adding synthetic samples declines obviously, except for RF. And the standard deviation of OKL just keeps around the one with no synthetic samples added. These trends are almost consistent to Fig. 11(a). And in Table 8, the standard deviation values of all methods fluctuate around the

TABLE 8. Standard deviation of 30 trials with different number of synthetic samples added in Test B.

Standard deviation	0	100	200	300	400	500
OKL	0.004	0.0038	0.005	0.0052	0.0041	0.0047
SVM	0.0059	0.0065	0.0043	0.0045	0.0054	0.0056
RF	0.0091	0.0091	0.0115	0.0065	0.0083	0.008
SBELM	0.9393	1.4887	1.1963	1.2707	0.7786	0.7999
ELM	0.019	0.0126	0.0142	0.0153	0.0212	0.0261

first column except for SBELM which gets drastic climbing. And we notice another unexpected result about RF which actually goes down when more synthetic samples are added. To testify the confidence about these comparative results, we further introduce Rank-sum Test to calculate the statistical significance between the results before and after adding synthetic samples. The p-values of Rank-sum Test are listed in Table 9 and 10.

TABLE 9. The p-values of Rank-sum Test in Test A.

p-value	100	200	300	400	500
OKL	0.248	5.43E-04	2.66E-05	2.64E-04	3.11E-06
SVM	2.22E-07	1.20E-04	2.49E-04	3.62E-04	0.0169
RF	3.05E-08	1.19E-08	2.23E-10	1.04E-10	1.87E-10
SBELM	0.0094	1.08E-06	7.80E-10	7.84E-11	3.51E-11
ELM	9.23E-11	5.35E-11	5.84E-11	6.25E-11	6.41E-11

TABLE 10. The p-values of Rank-sum Test in Test B.

p-value	100	200	300	400	500
OKL	0.2378	0.586	0.0598	1.02E-04	7.09E-09
SVM	3.26E-05	1.98E-05	2.13E-05	1.84E-07	2.66E-07
RF	0.6092	0.9232	0.818	0.5776	0.5383
SBELM	2.92E-11	2.91E-11	2.93E-11	2.93E-11	2.92E-11
ELM	3.22E-06	0.0554	8.58E-06	4.98E-04	9.26E-04

Obviously, the p-values in Table 9 are almost less than 0.05, which indicates the existence of significant difference between the standard deviation values before and after adding the spectral synthesis sample. And in Table 10, OKL and RF both have p-values more than 0.05, which indicates in turn that the results of OKL and RF in Table 8 are of no statistical confidence.

Therefore, considering the test accuracy in Table 5 and 6, it can be seen that the synthetic samples generated by GAN, no matter the synthetic spectrum samples or the feature samples, are helpful for fault diagnosis in most cases. But when further considering the stability, we can conclude that the frequency spectrum data are obviously superior to the 59-dimensional feature data in the generation of synthetic sample by GAN.

What needs to be supplied here is that, the time-domain vibration signal is another solution for generating synthetic samples by GAN. Here we also provide a simple evaluation for it. Taking ball-1 fault type as example, the synthetic samples generated from time-domain signal by GAN is shown in Fig. 12. It is clear that too much noise arouse in the

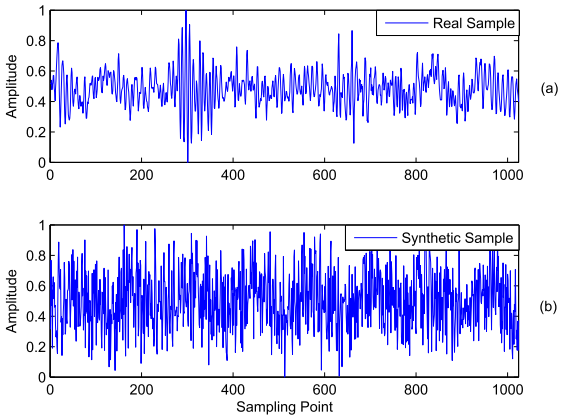

FIGURE 12. Comparison of (a) original time-domain signal and (b) synthetic sample by GAN for ball-1 fault type.

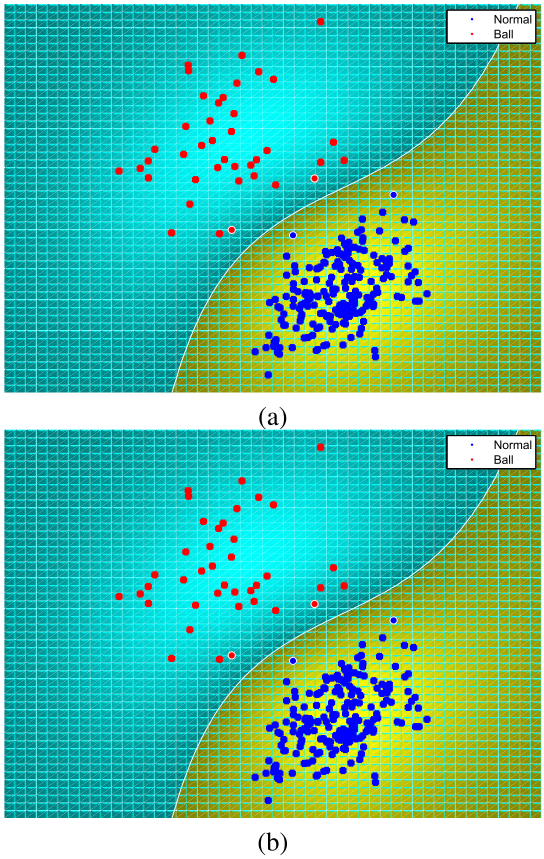

FIGURE 13. SVM hyperplanes between normal state and ball fault on (a) balanced data set and (b) imbalanced data set with ratio 5:1.

synthetic samples, which indicates that GAN can't learn well the data distribution of the time-domain vibration signal. There are two reasons accounting for this. One is that the original vibration signal indeed contains sort of noise, which will disturb the true data distribution and then cause decline of GAN's learning performance. Compared to time-domain signal, frequency spectrum obtained by FFT has strong regularity and can make quantitative interpretation of the vibration signal. The other one is that the original vibration signal

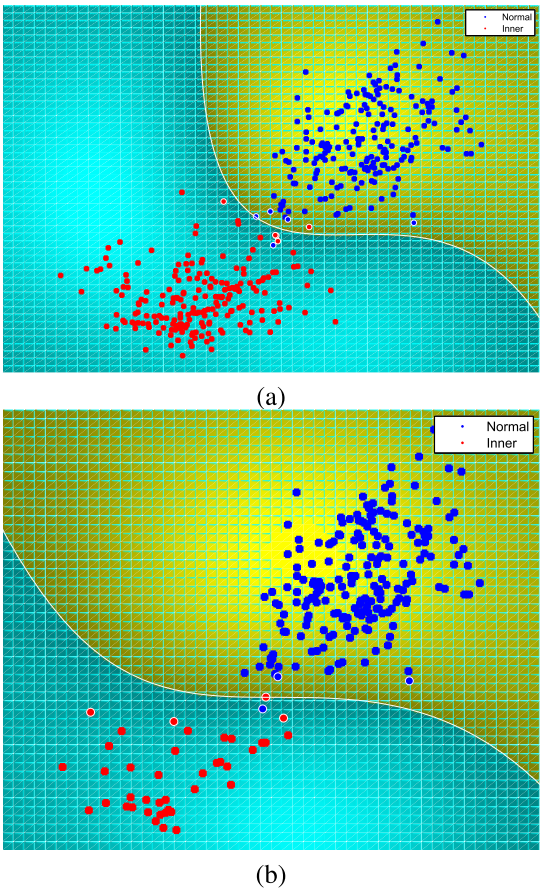

FIGURE 14. SVM hyperplanes between normal state and inner race fault on (a) balanced data set and (b) imbalanced data set with ratio 5:1.

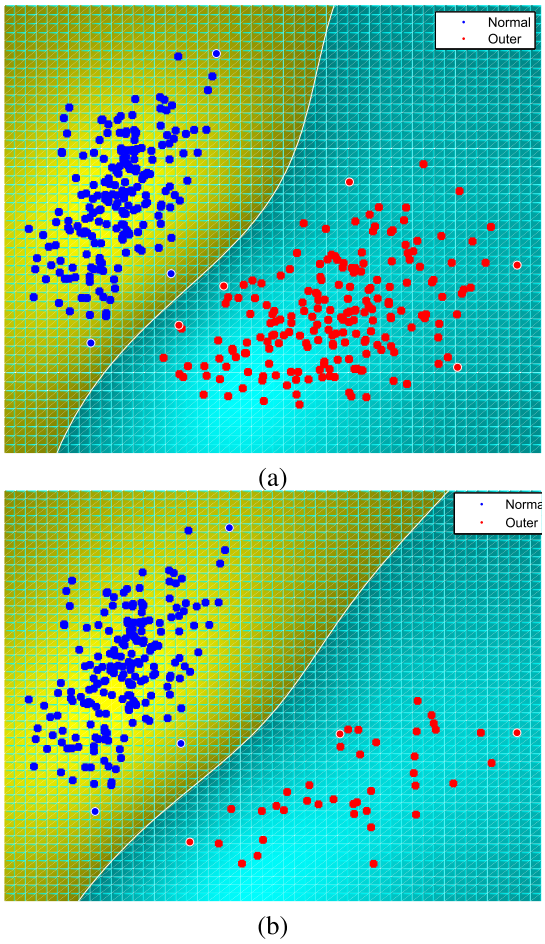

FIGURE 15. SVM hyperplanes between normal state and outer race fault on (a) balanced data set and (b) imbalanced data set with ratio 5:1.

is a time series with coupling between successive sampling points, and in this scenario, GAN couldn't learn the coupling relationship well.

In conclusion, using the frequency spectrum to generate synthetic samples by GAN, we can supplement the training set and improve the classification performance. Thus, we select spectrum synthetic samples by GAN to improve the diagnosis performance on imbalanced fault data in Experiment 2.

C. EXPERIMENT 2

In actual applications, the collected rolling bearing data is generally imbalanced, which means the amount of normal data is far more than that of fault data. In this case, the classification accuracy of traditional fault diagnosis methods is easy to be limited. Here we give some visual explanation of negative effect of imbalanced fault data, as illustrated in Fig. 13-15. We choose 500 normal samples and 500 fault samples for each fault type as well. Then we can establish three SVM classification models between normal and three fault classes respectively. And to provide imbalance effect, we shrink the fault samples to 20% of original scale through random under-sampling. It's clear that the SVM hyperplane tends to move to the ball class (minority) in Fig. 13(b), which will cause the test sample to be recognized as normal class (majority) more easily. Same effects, also called model bias, can be found in Fig. 14 and 15.

From Experiment 1, we know that GAN can generate useful synthetic samples to improve the generalization ability of classification model. In Experiment 2, we use the spectrum synthetic samples to balance the training set and run SDAE to complete feature extraction and final diagnosis. The target is to improve the minority test accuracy as well as the whole test accuracy and other evaluation indexes in imbalanced fault diagnosis problem.

1) EXPERIMENTAL SETTING

As a comparative study, this experiment not only compares with the diagnostic effect of SVM, RF, ELM, SBELM and OKL, but also compares with some widely-used oversampling methods such as Random Oversampling (RO) technology [31], SMOTE [32] and the Principal Curve-based (PC) method [5]. For these five classification models, spectrum synthetic samples by GAN are firstly generated to balance the training data, and time-domain statistical feature as listed in Table 3 are chosen to construct the training set

TABLE 11. The sample information used in Experiment 2.

Type	Load	Damage radius(inches)	Number
Normal	1	0.007	200
Normal	1	0.014	200
Normal	1	0.021	200
Ball	1	0.007	200
Ball	1	0.014	200
Ball	1	0.021	200
Inner race	2	0.007	200
Inner race	2	0.014	200
Inner race	2	0.021	200
Outer race	1	0.007	200
Outer race	1	0.014	200
Outer race	1	0.021	200

TABLE 12. The various imbalance ratios of the training set.

Normal	500	500	500	500	500	500	500
Ball	5	10	25	50	100	250	500
Inner race	5	10	25	50	100	250	500
Outer race	5	10	25	50	100	250	500
Imbalance ratio	100:1	50:1	20:1	10:1	5:1	2:1	1:1

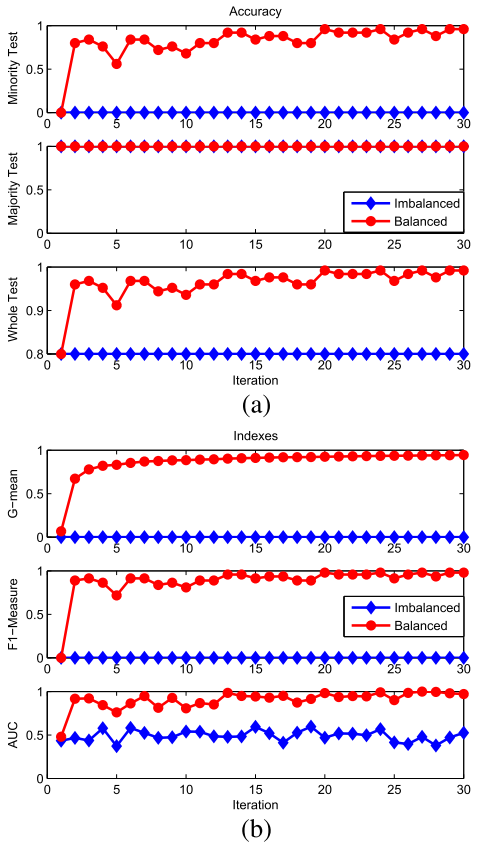

FIGURE 16. Comparative results of the proposed method for normal vs. ball fault in terms of (a) Accuracy and (b) three evaluation indexes.

for model training. For short, we call directly these five models by their name. The proposed method is abbreviated as GAN_SDAE.

We conduct 3 kinds of imbalanced bi-classification experiment on CWRU dataset: normal state vs. ball fault, normal state vs. inner race fault, normal state vs. outer race fault.

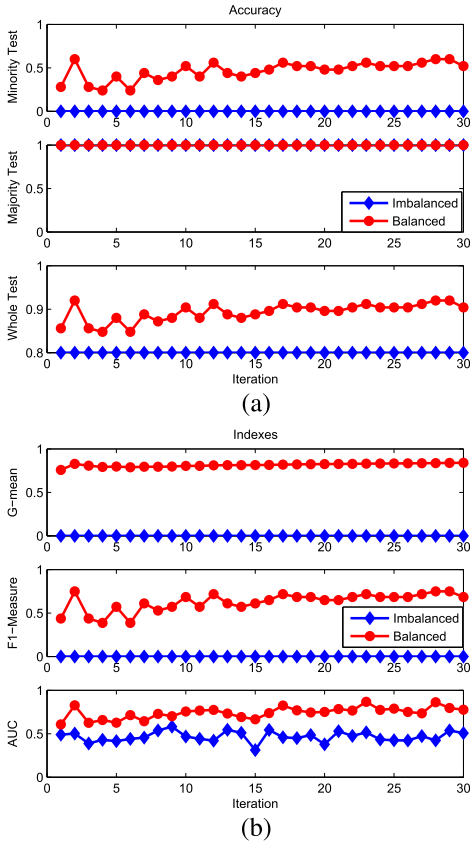

FIGURE 17. Comparative results of the proposed method for normal vs. inner race fault in terms of (a) Accuracy and (b) three evaluation indexes.

The sample information is shown in Table 11. We randomly select 500 samples of each class for training and the rest for test. To evaluate the effectiveness of the proposed method, we set various imbalance ratios on the training set in these experiments, as listed in Table 12.

To make a comprehensive comparison, we adopt five evaluation indexes: the minority test accuracy, the whole test accuracy, F1_Measure [33], G_mean [5] value and AUC value. The AUC value is defined as the area of the ROC curve.

Among them we mainly focus on the minority test accuracy and F1_Measure. The larger values of these two indexes mean better diagnosis performance.

2) RESULT AND ANALYSIS

Due to space limitation, this experiment only takes the values before and after balance for comparison, with the imbalance ratio 50:1 as a representative. Figure 16-18 show the comparative results of the proposed method for normal vs. ball fault, normal vs. inner race fault and normal vs. outer race fault, respectively. The iteration number of SDAE is set 30. And the parameter settings of SDAE in Experiment 2 is same as that of in Experiment 1.

From Fig. 16-18, the proposed method has significant improvement in terms of all evaluation indexes, indicating that it is effective to use GAN to generate minority class

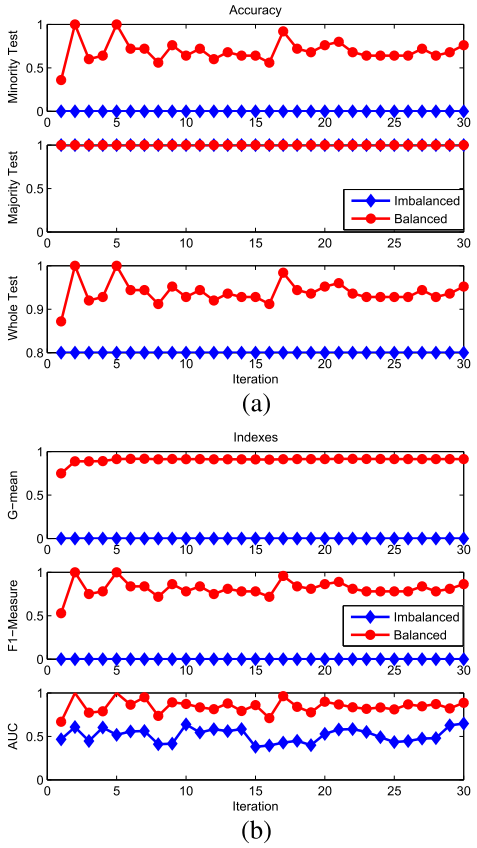

FIGURE 18. Comparative results of the proposed method for normal vs. outer race fault in terms of (a) Accuracy and (b) three evaluation indexes.

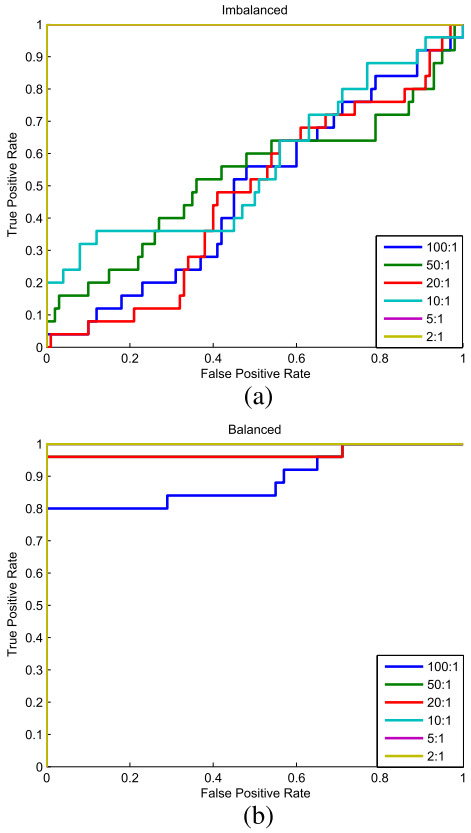

FIGURE 19. ROC curves on normal and ball fault dataset at different imbalanced ratios: (a) the original dataset and (b) add the GAN synthetic samples.

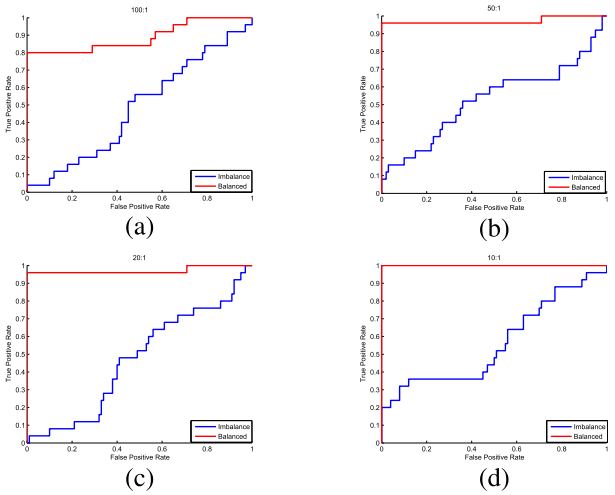

FIGURE 20. ROC curves before and after balance with different imbalance ratios at (a) 100:1, (b) 50:1, (c) 20:1, (d) 10:1.

synthetic samples for solving the problem of imbalanced fault diagnosis.

Besides the indexes mentioned above, we also adopt ROC curves, which is usually used to evaluate the imbalanced classification performance due to its invariant property in the distribution transformation of positive and negative samples. Here, we only give the ROC curves of the proposed method on the normal samples and ball fault samples in Fig. 19 with different imbalance ratios.

In Fig. 19, we observe an interesting phenomenon. With the ratio 2:1, the ROC curve rises up quickly, no matter on imbalanced data or balanced set. But with the ratio 100:1, the blue line in Fig. 19(b) has a slow ascending trend, even if GAN is introduced to generate synthetic samples which are supposed to have as similar distribution as possible with the original signal. That indicates that in the severe imbalance problem, the minority sample could not provide enough prior information about data distribution for GAN. This problem should be considered more in actual applications.

Moreover, some ROC curves in Fig. 19(b) are overlapped. In order to make a clearer comparison, we bring out the ROC curves with different imbalance ratios from Fig. 19, and show them separately in Fig. 20. It can be seen clearly that the proposed method can improve the performance largely on the imbalanced fault diagnosis problem.

As a comparative study, this experiment also compares the proposed method with SVM, ELM and OKL etc. To make the results stable, all models for comparison take the mean value of 30 repeated trials as the final result. For intuitive understanding, Fig. 21 shows the graphical comparison with the imbalance ratio fixed as 50:1. Here we only display the

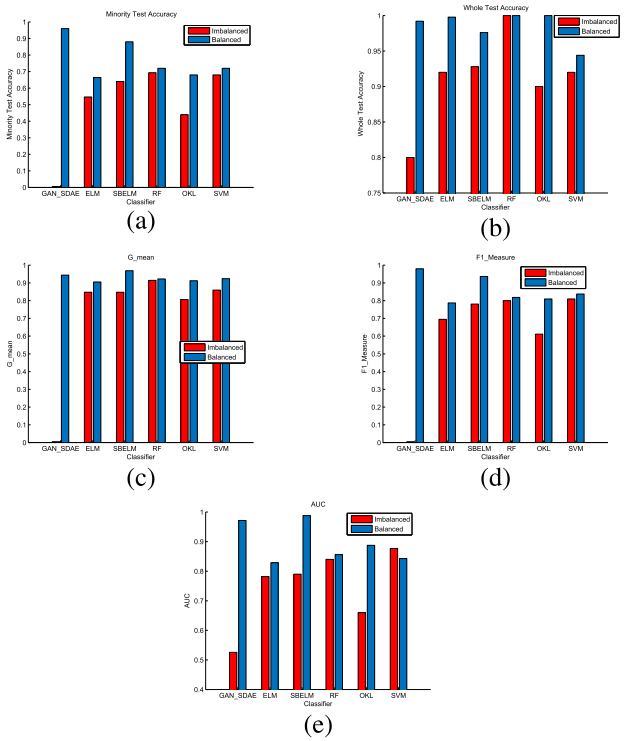

FIGURE 21. **Comparative results of different models with various imbalance ratios in terms of (a) Minority test accuracy, (b) whole test accuracy, (c) G_mean, (d) F1_Measure, (e) AUC.**

diagnosis result for normal vs. ball fault. Similar comparison can be found for normal vs. inner/outer race fault.

From Fig. 21, the proposed method gets better overall performance than other five models. Especially in terms of minority test accuracy and F1_Measure which two are most valued, the proposed method obtains best result. And the other five models all get improvement of prediction performance except of SVM in terms of AUC, which demonstrates again the effectiveness of GAN for oversampling on imbalanced fault diagnosis problem.

Another question is the very low performance of SDAE before balance as shown in Fig. 21, no matter which index adopted. In this scenario, SDAE is directly used to extract deep features from original spectrum data. However, as the imbalance effect is severe, the minority sample wouldn't provide too much useful information for feature extraction. Although the statistical features used for the other five models look rather simpler than deep feature, they work effectively on minority samples. Of course, once the training set is balanced, deep features start to show their promising performance, even extracted from synthetic samples.

We further verify the performance of different methods with various imbalance ratios, as shown in Fig. 22. Along with the imbalance level going down, almost all methods get better diagnosis performance. Obviously, with all imbalance ratios the proposed method outperforms other five models in terms of minority test accuracy and F1_Measure.

Fig. 23 shows the ROC curve of different models with the imbalance ratio 50:1. Same to Fig. 21, the proposed method

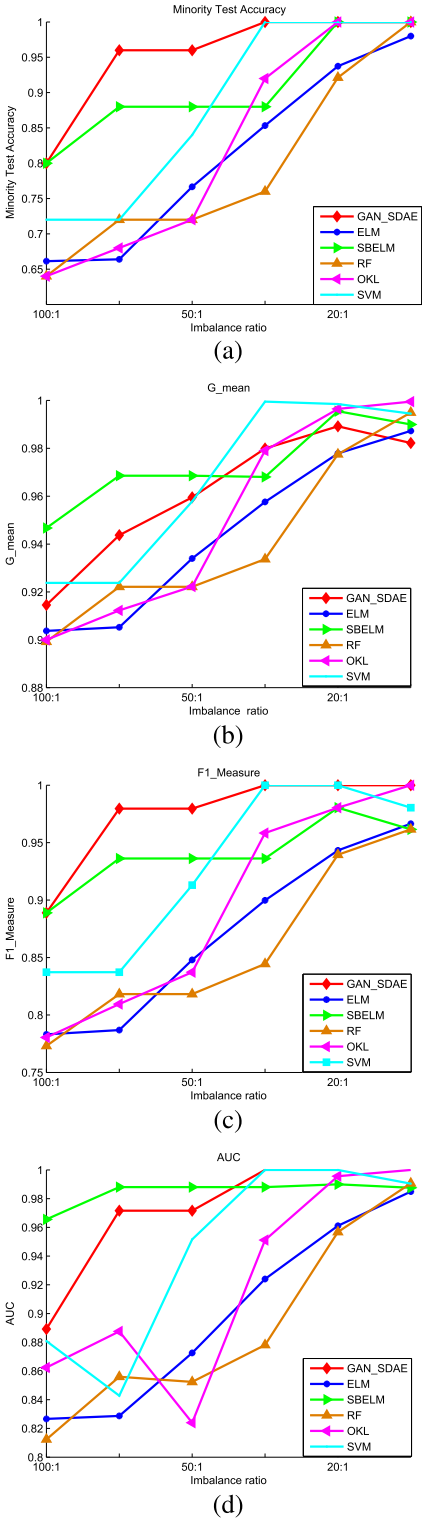

FIGURE 22. **Performance of different methods with various imbalance ratios in terms of (a) Minority test accuracies, (b) G_mean, (c) F1_Measure, (d) AUC.**

gets worst performance before balance, as show in Fig. 23(a), and has a largest improvement in terms of AUC after balance. This comparison testifies again the superior performance of the proposed method.

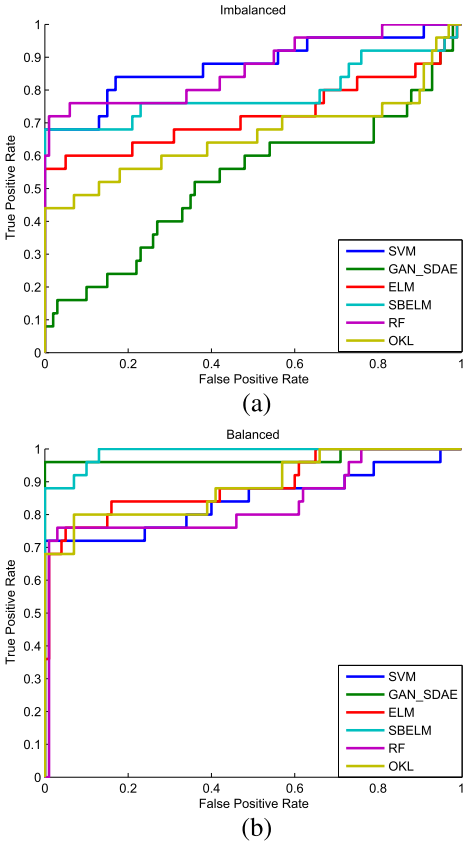

FIGURE 23. ROC curves of different models balance with imbalance ratios 50:1 on (a) the original imbalanced data and (b) balanced data with synthetic samples.

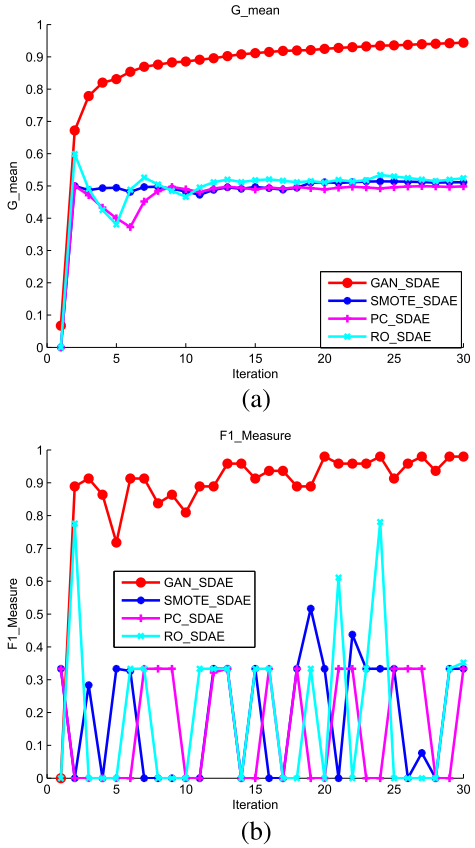

FIGURE 24. Comparison of the proposed method and three oversampeling techniques with SDAE in terms of (a) G_mean and (b) F1_Measure.

As mentioned above, we also plan to compare with three widely-used oversampling algorithms: RO, SMOTE and PC. We integrate these three algorithms with different classification models. For convenience, we abbreviate this kind of combination method as RO_SDAE, SMOTE_SDAE and PC_SDAE etc. The imbalance ratio is still fixed as 50:1. Fig. 24 provides the G_mean and F1_Measure of these four oversampling methods with SDAE for normal vs. ball fault. Without doubt, the proposed method GAN_SDAE gets best performance. More importantly, from the curve trend in Fig. 24, the proposed method is more stable than the other three methods.

Moreover, we try to make a comprehensive comparison by integrating these three oversampling techniques with different classification models. Due to space limitation, we only choose RF and SBELM as representative. Considering the combination of two methods, we still adopt the naming rule mentioned above. The comparative results for normal vs. ball fault are listed in Table 13. The imbalance ratio is set 50:1. The final results of RF and SBELM are the mean value of 30 trials.

From Table 13, the proposed method outperforms other methods in terms of minority test accuracy and F1_Measure. And with RF, SBELM and SDAE, GAN all gets much better performance than other three oversampling techniques.

TABLE 13. Comparative results of different methods for normal vs. ball fault.

Method	Minority Test Accuracy	Minority Training Accuracy	Whole Test Accuracy	G_mean	F1_Measure	AUC
RO_RF	0.6747	0.99	1	0.9086	0.7869	0.8282
SMOTE_RF	0.72	0.9703	1	0.9179	0.7832	0.8417
PC_RF	0.688	0.99	1	0.9126	0.7962	0.8468
GAN_RF	0.72	0.99	1	0.9221	0.8182	0.988
RO_ SBELM	0.76	1	0.952	0.9359	0.8636	0.8724
SMOTE_SBELM	0.84	0.984	0.944	0.9458	0.8571	0.9224
PC_ SBELM	0.88	0.996	0.976	0.9675	0.9362	0.9324
GAN_ SBELM	0.88	0.998	0.976	0.9685	0.9362	0.988
GAN_SDAE	**0.96**	**1**	0.992	0.9437	**0.9796**	0.9716

Therefore, we think that the proposed method is good at oversampling for minority class sample as well as the feature extraction for fault diagnosis.

V. CONCLUSION

In this paper a new imbalance fault diagnosis method is proposed based on GAN and SDAE. The method combines the advantage of GAN which can learn the data distribution automatically and SDAE which is able to extract deep feature adaptively. For this comparative study, we can draw some conclusions from the experimental results on CWRU dataset:

(1) Compared to the widely-used oversampling techniques, the proposed method can generate synthetic samples with better quality which is helpful to improve diagnosis performance. And the frequency spectrum of original signal is preferable for GAN-based fault diagnosis. (2) Compared to the other shallow classification models, the proposed method always performs better in terms of minority test accuracy and F1_Measure. (3) For severe class imbalance problem,

only using GAN is hard to generate satisfied enough synthetic samples due to lack of prior information. In this scenario, quite different from traditional statistical features, deep model such as SDAE can make good use of such synthetic samples and then extract rich and representative features for final diagnosis.

In the following work, we will focus on exploiting distribution for flow data which locates in the bearing degradation process. This exploiting technique should consider more about the online data variation and conduct self-adaptive generation of synthetic sample. Anomaly detection using GAN is another interesting problem.

REFERENCES

[1] T.-Y. Liu and G.-Z. Li, "The imbalanced data problem in the fault diagnosis of rolling bearing," *Comput. Eng. Sci.*, vol. 32, no. 5, pp. 150–153, 2010.

[2] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[3] E. Ramentol, Y. Caballero, R. Bello, and F. Herrera, "SMOTE-RSB: A hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory," *Knowl. Inf. Syst.*, vol. 33, no. 2, pp. 245–265, 2012.

[4] M. Gao, X. Hong, S. Chen, and C. J. Harris, "A combined SMOTE and PSO based RBF classifier for two-class imbalanced problems," *Neurocomputing*, vol. 74, no. 17, pp. 3456–3466, 2011.

[5] W. Mao, L. He, Y. Yan, and J. Wang, "Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine," *Mech. Syst. Signal Process.*, vol. 83, pp. 450–473, Jan. 2017.

[6] 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. Signal Process.*, vol. 110, pp. 349–367, Sep. 2018.

[7] J.-M. Yin, M. Yang, and J.-W. Wan, "A kernel Fisher linear discriminant analysis approach aiming at imbalanced data set," *Pattern Recognit. Artif. Intell.*, vol. 23, no. 3, pp. 414–420, 2010.

[8] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognit.*, vol. 40, no. 12, pp. 3358–3378, 2007.

[9] H. Xiang, Y. Yang, and S. Zhao, "Local clustering ensemble learning method based on improved AdaBoost for rare class analysis," *J. Comput. Inf. Syst.*, vol. 8, no. 4, pp. 1783–1790, 2012.

[10] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, 2014, pp. 2672–2680.

[11] X. Liang, Z. Hu, H. Zhang, C. Gan, and E. P. Xing. (2017). "Recurrent topic-transition GAN for visual paragraph generation." [Online]. Available: https://arxiv.org/abs/1703.07022

[12] W. Fedus, I. Goodfellow, and A. M. Dai. (2018). "MaskGAN: Better text generation via filling in the _____." [Online]. Available: https://arxiv.org/abs/1801.07736

[13] A. Radford, L. Metz, and S. Chintala. (2015). "Unsupervised representation learning with deep convolutional generative adversarial networks." [Online]. Available: https://arxiv.org/abs/1511.06434

[14] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," ICML, Sydney, NSW, Australia, Tech. Rep., 2017.

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

[16] Y. O. Lee, J. Jo, and J. Hwang, "Application of deep neural network and generative adversarial network to industrial maintenance: A case study of induction motor fault detection," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2017, pp. 3248–3253.

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

[18] T. Han, C. Liu, W. Yang, and D. Jiang, "A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults," *Knowl.-Based Syst.*, vol. 165, pp. 474–487, Feb. 2019, doi: 10.1016/j.knosys.2018.12.019.

[19] P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.

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

[21] *These Data Comes From Case Western Reserve University Bearing Data Center*. Accessed: Dec. 12, 2018. [Online]. Available: http://www.eecs.cwru.edu/laboratory/bearings/

[22] M. van Heeswijk and Y. Miche, "Binary/ternary extreme learning machines," *Neurocomputing* vol. 149, pp. 187–197, Feb. 2015.

[23] J. Luo, C.-M. Vong, and P.-K. Wong, "Sparse Bayesian extreme learning machine for multi-classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 4, pp. 836–843, Apr. 2014.

[24] J.-S. Wong and M. J. Lancaster, "Microstrip filters for RF/microwave applications [book review]," *IEEE Microw. Mag.*, vol. 3, no. 3, pp. 62–65, Sep. 2002.

[25] F. Dinuzzo, C. S. Ong, P. Gehler, and G. Pillonetto, "Learning output kernels with block coordinate descent," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 49–56.

[26] W. T. Mao, M. Tian, and G. R. Yan, "Research of load identification based on multiple-input multiple-output SVM model selection," *Proc. Inst. Mech. Eng. C, J. Mech. Eng. Sci.*, vol. 226, no. 5, pp. 1395–1409, 2012.

[27] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu, "Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing," *Mech. Syst. Signal Process.*, vol. 100 pp. 743–765, Feb. 2018.

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

[29] T. W. Rauber, F. de A. Boldt, and F. M. Varejão, "Heterogeneous feature models and feature selection applied to bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 637–646, Jan. 2015.

[30] N. Pezzotti, B. P. F. Lelieveldt, L. van der Maaten, T. Höllt, and E. Eisemann, "Approximated and user steerable tSNE for progressive visual analytics," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 7, pp. 1739–1752, Jul. 2017.

[31] H. Zhang and M. Li, "RWO-Sampling: A random walk over-sampling approach to imbalanced data classification," *Inf. Fusion*, vol. 20, no. 1, pp. 99–116, 2014.

[32] Z.-Q. Zeng, Q. Wu, and B.-S. Liao, "A classification method for imbalance data set based on kernel SMOTE," *Acta Electron. Sinica*, vol. 37, no. 11, pp. 2489–2495, 2009.

[33] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, "Optimal thresholding of classifiers to maximize F1 measure," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2014, pp. 225–239.

WENTAO MAO received the M.S. degree in computer science from the Chongqing University of Posts and Telecommunications, in 2006, and the Ph.D. degree in engineering mechanics from Xi'an Jiaotong University, China, in 2011. He is currently with Henan Normal University, China. He is involved in about 10 research projects, such as the National Natural Science Foundation of China project as a Project Principal or a Main Researcher. His current research interests include fault diagnostics and prognostics, machine learning, and data mining.

YAMIN LIU received the bachelor's degree in computer science and technology from Henan Normal University, China, in 2018, where she is currently pursuing the M.S. degree. Her research interests include deep learning and big data analysis.

LING DING received the bachelor's degree in computer science and technology from Henan Normal University, China, in 2018, where she is currently pursuing the M.S. degree. Her research interests include deep learning and fault diagnosis.

YUAN LI received the bachelor's degree from Nanjing Normal University, China, in 2010, and the Ph.D. degree in mechanical engineering from Hunan University, China, in 2015. She is currently an Associate Professor with Henan Normal University, China. Her research interests include early fault diagnosis and reliability analysis.