Enhanced generative adversarial networks for bearing imbalanced fault diagnosis of rotating machinery

Yandong Hou[1] · Jiulong Ma[1] · Jinjin Wang[1] · Tianzhi Li[1] · Zhengquan Chen[2]

Abstract

Traditional rolling bearing fault diagnosis approaches require a large amount of fault data in advance, while some specific fault data is difficult to obtain in engineering scenarios. This imbalanced fault data problem seriously affects the accuracy of fault diagnosis. To improve the accuracy under imbalanced data conditions, we propose a novel data augmentation method of Enhanced Generative Adversarial Networks with Data Selection Module (EGAN-DSM). Firstly, a network enhancement module is designed, which quantifies antagonism between the generator and discriminator through loss value. And the module determines whether to iteratively enhance the networks with weak adversarial ability. Secondly, a Data Selected Module (DSM) is constructed using Hilbert space distance for screening generated data, and the screened data is mixed with original imbalanced data to reconstruct balanced data sets. Then, Deep Convolutional Neural Networks with Wide First-layer Kernels (WDCNN) is used for fault diagnosis. Finally, the method is verified by data measured on a rotating machine experimental platform. The results show that our method has high fault diagnosis accuracy under the condition of imbalanced data.

Keywords Data augmentation · Data selection module · Generative adversarial networks · Imbalanced data · Fault diagnosis

1 Introduction

Bearings are an important supporting component in industrial equipment, and their health state will affect working performance of the entire equipment. Therefore, it is of great significance to timely diagnose bearings faults to main-

Yandong Hou, Jiulong Ma and Zhengquan Chen contributed equally to this work.

✉ Zhengquan Chen
zqchen@vip.henu.edu.cn

Yandong Hou
hydong@henu.edu.cn

Jiulong Ma
mjl@henu.edu.cn

Jinjin Wang
wangjinjin@henu.edu.cn

Tianzhi Li
litianzhi@henu.edu.cn

[1] School of Artificial Intelligence, Henan University, 450046 Zhengzhou, China

[2] School of Computer and Information Engineering, Henan University, 475004 Kaifeng, China

tain safe equipment operation [1–3]. Most classical bearing fault diagnosis methods are based on the decomposition and transformation of fault signals, such as Empirical Mode Decomposition (EMD) [4], and Wavelet Transform (WT) [5]. But in recent years, deep learning has developed rapidly in the field of bearing fault diagnosis with its powerful deep feature extraction ability [6–10]. Deep learning models such as Convolutional Neural Networks (CNN) [11], Deep Belief Networks (DBN) [12], and Stacked Auto-Encoder (SAE) [13] have achieved great success in the field of bearing fault diagnosis as combined with vibration signal preprocessing [14–16]. However, a large amount of fault bearing data is required for these methods to achieve better fault diagnosis result[17, 18]. But bearings work normally in actual process of production, resulting in fault data being significantly less than normal. The problem of data imbalance will seriously affect the diagnostic performance of deep learning model [19–21].

The current research methods to solve data imbalanced problem can be divided into algorithm optimization and data expansion. Algorithm optimization pays more attention to minority class data by optimizing structure or loss function of the diagnostic model. However, optimization-based methods finitely improve diagnostic performance when minority

class data is extremely sparse [22]. Data expansion augments imbalanced data by synthesizing minority class data [23]. Chawla et al. [24] proposed the classical Synthetic Minority Over-Sampling Technique (SMOTE), which generates new samples by random linear interpolation among the minority class of samples. To increase the influence of data on the boundary of distribution, Han et al. [25] divided minority data into three classes and proposed Borderline-SMOTE. However, some minority class may also be difficult to classify. Therefore, He et al. [26] proposed Adaptive Synthetic Sampling (ADASYN) to adaptive put classification decision boundary toward hard identification data.

Although above oversampling methods can expand the data of minority class, oversampling method relies on data features. These methods rarely consider real distribution of minority class data and the generated data with less potential feature information [27]. Therefore, the method of data generation is proposed to deeply mine feature information of minority class data. Data generation generates data by learning the distribution of data [28], and its major research methods are Variational Auto-Encoder (VAE) [29] and Generative Adversarial Networks (GAN) [30]. Zhao et al. [31] achieved good results by generating minority class data through VAE to supplement imbalanced data. However, VAE is an explicit density model, which limits the powerful fitting ability of neural networks and makes data generation inefficient [32]. Therefore, GAN with implicit density models has been developed to adequately use the fitting ability of neural networks.

Recently, GAN-based fault diagnosis methods for imbalanced data have been explored preliminarily [33]. To solve the problem of imbalanced data in engineering scenarios, Shao et al. [34] adopted the Auxiliary Classifier Generative Adversarial Networks (ACGAN) model to augment induction motor vibration signal datasets. By overcoming mode collapse and vanishing gradient, Li et al. [27] improved the ACGAN model for quality generated data which can result in better diagnosis accuracy. Although these GAN-based methods are able to address the problem of data imbalance, the GAN still suffers from training instability due to the presence of adversarial properties [35]. The excessive learning ability of sub-neural networks can inhibit learning ability of the other networks, resulting in generated data mismatching the distribution of real data [27]. Therefore, in order to alleviate the learning inhibition between sub-networks, an EGAN based on Wasserstein Generative Adversarial Networks-Gradient Penalty (WGAN-GP) is designed in this paper. The network not only inherits the advantages of WGAN-GP in resolving gradient explosion and gradient disappearance during training [36] but can also adjust the learning differences between sub-networks promptly.

In addition, the data generated by GAN will also change due to adversarial training. However, the method mentioned above does not effectively process the generated data [37]. We designed a data selected module with Maximum Mean Discrepancy (MMD) to filter all the data generated by EGAN due to the property that MMD can measure the similarity of two distributions [38]. In this paper, a data augmentation method of EGAN-DSM is proposed, and WDCNN is combined to classify bearing fault data. Experimental results show that the proposed method can improve diagnostic accuracy of imbalanced data sets. The larger imbalanced ratio, the more obvious improvement effect. The main contributions are shown as follows:

1. An enhanced module is developed to enhance adversarial ability of two sub neural networks. This module can solve low-quality generated data caused by the suppression of two sub-neural networks effectively.
2. DSM is designed as a post-processing module for the output data of GAN. DSM is used to filter generated data of GAN and find out data that contains the most feature information of real data.
3. An intelligent fault diagnosis method that solves imbalanced data effectively is designed: EGAN-DSM-WDCNN. Imbalanced training data sets are balanced by the method of EGAN, and then the balanced data is used to train WDCNN. This method can improve diagnostic accuracy effectively.

The rest of this article is arranged as follows. Section 2 introduces theoretical background of the proposed method in detail, including relevant knowledge of WGAN-GP and WDCNN. Section 3 presents details of the proposed approach. Section 4 shows simulation results of the proposed method. Section 5 presents the conclusion of this paper.

2 Theoretical background

The overall framework of the proposed method is divided into four stages: data preprocessing, data generation, data selection, and fault classification. WGAN-GP is prototyped for data generation and WDCNN is chosen for fault classification stage. Theoretical background of the used methods will be briefly described in this section.

2.1 GAN

GAN consists of two sub-neural networks, one called generator and the other called discriminator. These two sub-neural networks are trained against to optimize the output of GAN according to the idea of zero-sum game. Its value function is shown in Eq. (1):

$$\Psi(G, D) = E_{x \sim P_x}[\log(D(x)] + E_{N \sim P_N}(\log(1 - D(q)) \quad (1)$$

where $E_{x \sim P_x}$ represents the expectation of real data x under distribution P_x. $E_{N \sim P_N}$ is expectation of noise N under its prior distribution P_N. G represents generator, D represents discriminator. The input of generator is N, and the output is $q = G(N)$. $D(q)$ is output of discriminator when its input is q. Generator generates data with same probability distribution as original input. Discriminator determines the distribution gap between generated data and original data. The two expectation sums $\Psi(G, D)$ in Eq. (1) represents Jensen-Shannon divergence of the distribution of generated data and original data. The training strategy of whole networks is shown in Eq. (2):

$$\min_G \max_D \Psi(G, D) \tag{2}$$

when generator is fixed, discriminator is trained by maximizing $\Psi(G, D)$ to distinguish original data and generated data. Then, discriminator is fixed and generator is trained by minimizing $\Psi(G, D)$ to make generated data as realistic as possible. Both discriminator and generator could reach a local optimum after extensive adversarial training. It means that data generated by generator has the same distribution as original data and discriminator cannot distinguish generated data from original data.

2.2 Wasserstein distance

GAN evaluates the distribution of generated data and original data through Jensen-Shannon divergence, and then optimizes it. However, the discriminator may reach saturation in advance during training because of dispersion of Jensen-Shannon divergence. That results in disappearance of gradient during training [39]. In order to solve this problem, Arjovsk [40] replaced Jensen-Shannon divergence with wasserstein distance, as shown in Eq. (3):

$$W(P_x, P_q) = \inf_{\eta \sim \Gamma(P_x, P_q)} E_{(x,q) \sim \eta}[\|x - q\|] \tag{3}$$

where $W(P_x, P_q)$ represents wasserstein distance of generated data distribution P_q and original data distribution P_x. $\Gamma(P_q, P_x)$ represents a set of all joint distributions with P_x and P_q. η is any possible joint distribution of $\Gamma(P_q, P_x)$, and (x, q) is sampled from η. $\inf\{\cdot\}$ represents the infimum. According to Kantorovich Rubinstein duality theorem, the calculation of $W(P_x, P_q)$ can be transformed into Eq. (4).

$$W(P_x, P_q) = \sup_{\|f\|_L \leq 1} (E_{x \sim P_x}[f(x)] - E_{N \sim P_N}[f(q)]) \tag{4}$$

where $\sup\{\cdot\}$ represents supremum. f represents a 1-Lipschitz continuous function, and satisfies Eq. (5).

$$\|f\|_L \triangleq \sup_{z \neq e} \frac{f(z) - f(e)}{z - e} \leq 1 \tag{5}$$

Value function of WGAN can be obtained by combining the above functions as follows.

$$\min_G \max_{D \in \Omega} \left\{ E_{x \sim P_x}[D(x)] - E_{N \sim P_N}[D(q)] \right\} \tag{6}$$

where Ω denotes a set of 1-Lipschitz functions. WGAN limits the weight of discriminator to $[-\omega, \omega]$ by gradient clipping strategy. The benefit of this way is that the networks training is no longer limited by dispersion of Jensen-Shannon divergence, which improves stability of networks training.

2.3 Gradient Penalty

Wasserstein distance is developed to improve stability of whole networks. Nevertheless, the networks still suffer problems of being unable to converge and low quality of generated data. Due to using weight clipping to satisfy Lipschitz constraints, WGAN works as a simple function and fails to fit some complex data. Therefore, Gulrajani et al. [41] proposed WGAN-GP by replacing the weight clipping of WGAN with gradient penalty. Its value function is shown in Eq. (7):

$$E_{x \sim P_x}[D(x)] - E_{N \sim P_N}[D(q)]$$
$$-\lambda E_{\hat{x} \sim P_{\hat{x}}}[(\left\|\nabla_{\hat{x}} D(\hat{x})\right\|_2 - 1)^2] \tag{7}$$

where the last term is penalty term, λ is penalty coefficient. $\hat{x} = \varepsilon x + (1 - \varepsilon)q$ represents samples on all straight lines between generated data distribution and original data distribution. In general, a 1-Lipschitz function refers to the gradient norm of a function is at most 1. However, the penalty term in Eq. (7) penalizes gradient norm away from 1, driving all norms toward 1. This method can stabilize the gradient and ensure that generated data distribution gradually gets closer to real data distribution.

2.4 WDCNN

Due to the excellent fitting effect of CNN, it is widely used in various scenarios, including bearing fault diagnosis. Samples would be long when a 1-dimensional bearing vibration signal sample contains enough data information. However, a convolutional model with 3×1 small convolutional kernels to learn that long vibration signals would make the training of model hard. Zhang et al. [42] designed WDCNN, which can capture useful information in middle and low-frequency vibration signals. Then successive 3×1 small kernel convolutional layers are used to obtain better feature representation.

The way to process data is as follows algorithm 1. Zhang's experimental results show that WDCNN has better diagnostic accuracy for bearing data.

Algorithm 1 WDCNN

Input: mixed balanced data sets X_{mix}, $(x^0 = \{x_1, ..., x_j\} \in X_{mix})$, initial parameters K_0, b_0, γ_0, β_0, pooling width W_p, m classes bearing healthy state

Output: fault diagnosis accuracy L

1: **for** $n = 1, 2, 3, 4, 5$ **do**
2:     Convolution calculate x_j for local features of the next layer $y_j{}^n = K_{n-1} * x_j{}^{n-1} + b_{n-1}$
3:     $\hat{y}_j^n = (y_j{}^n - E[y_j{}^n])/\sqrt{Var[y_j{}^n]}$
4:     Batch Normalization local feature $y_j{}^n$ for $y_{j,BN}^n = \gamma_{n-1}\hat{y}_j^n + \beta_{n-1}$
5:     The nonlinear expression of $y_{j,BN}^n$, $a_j^n = \max\{0, y_{j,BN}^n\}$
6:     $P_j^n = \max\limits_{(j-1)W_p+1 \leq t \leq jW_p}\{a^n(t)\}$
7:     The next convolution feature extraction $x_j^n \leftarrow P_j^n$
8: **end for**
9: The probability distribution of different bearing healthy state $q = e^{P_j{}^5}/\sum_k^m e^{P_k{}^5}$
10: Compared q with real bearing health condition, L is obtained

---

3 Proposed method

An intelligent fault diagnosis method of EGAN-DSM-WDCNN is proposed so as to solve the problem of low diagnostic accuracy caused by sparse fault data in actual industry. EGAN and DSM of this approach are our significant innovations. EGAN can generate the scarce data we need and prevent the problem of learning inhibition in the training process. DSM improves the quality of generated data by utilizing data post-processing. Therefore, DSM can effectively screen out the data similar to the original data distribution in EGAN's generated data.

3.1 EGAN

To solve the problem of bearing imbalanced data, we propose an Enhanced Generative Adversarial Networks. EGAN is an enhanced model based on WGAN-GP. As a result, EGAN not only inherits the advantages of WGAN-GP in resolving gradient explosion and gradient disappearance during training but can also balance sub-network learning differences promptly. Specifically, increasing antagonism between the generator and discriminator prevents one sub-networks from inhibiting the learning ability of the other.

An enhanced processing module is designed in EGAN at the end of each training epoch. This module can increase training times of weaker networks when there is a certain gap between the loss values of the generator and discriminator. Considering the excellent performance of CNN in data feature extraction, a generator, and a discriminator are constructed. The generator is constructed based on 1-D deconvolution neural networks. Furthermore, discriminator is constructed based on 1-D convolutional neural networks. The specific structure diagram is shown in Fig. 1, where (a) is discriminator, (b) is generator, and (c) is the proposed enhancement module.

Gaussian noise is put into generator, and generated data with a distribution close to original data is output. A single class fault data is put into discriminator for the purpose of judging generated data. Generated data is put into discriminator, and discriminant value $D(q)$ is output. The loss value of generator $\tilde{g}$ is numerically equal to $D(q)$. Similarly, discriminant value $D(x)$ of real data can be obtained by putting the real data into the discriminator. The loss value $\tilde{d}$ of discriminator is obtained by putting $D(q)$ and $D(x)$ into Eq. (7). The targeted enhancement is achieved by distinguishing difference in loss value between two sub-neural networks. The details of enhanced module are shown in Fig. 1 (c).

3.2 Data Selected Module

It is not easy to achieve theoretical Nash equilibrium for GAN training. Two sub-neural networks train against each other, resulting in different features in the generated data. Therefore, we design a post-processing data selection module through which we can achieve the purpose of screening and filtering similar data distribution.

In DSM, a metric function is needed to quantify the probability distribution distance between generated data and original data. In addition, the Maximum Mean Discrepancy (MMD) is often used to measure the difference between two distributions. In this paper, MMD is used to quantify the similarity between generated data distribution and original data distribution.

Assuming that the original data $x = \{x_1, ..., x_n\}$ with n samples and generated data $q = \{q_1, ..., q_m\}$ with m samples. Then MMD between original data and generated data is shown in Eq. (8).

$$MMD[F, P_x, P_q] = \sup_{\beta \in F}(E_{P_x}[h(x)] - E_{P_q}[h(q)]) \quad (8)$$

where F represents an arbitrary vector in unit sphere in the regenerated Hilbert space. $h(\bullet)$ is shown in Eq. (9), denotes scalar product of vectors β in regenerated Hilbert space and vectors $f(\alpha)$ in original space.

$$h(\alpha) = \langle \beta, f(\alpha) \rangle_H \quad (9)$$

Fig. 1 Structure of EGAN model

(a) Discriminator

(b) Generator

(c) Enhance Modul
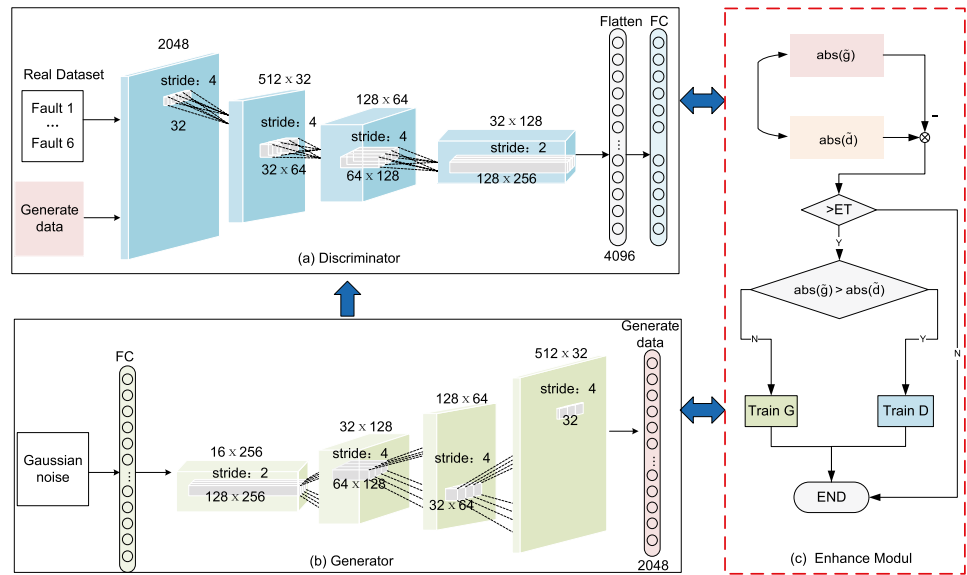
H denotes that data is mapped to Hilbert space through mapping function. According to Eq. (8) - Eq. (9) and [43], Eq. (10) can be deduced.

$$MMD[F, x, q] = \left\| \frac{1}{n} \sum_{i=1}^{n} f(x_i) - \frac{1}{m} \sum_{j}^{m} f(q_j) \right\|_H \quad (10)$$

DSM is used to map all data into Hilbert space to calculate the difference. Then the mean discrepancy between generated data and original data is calculated. And the maximum mean discrepancy is considered as the difference among these data. DSM selects a group of generated data with the smallest MMD as a complement of minority class data to achieve the purpose of balancing data.

3.3 Fault diagnosis method

The proposed fault diagnosis method can be divided into four stages, as shown in Fig. 2. Data preprocessing is called stage 1, where vibration signals of bearings are obtained from the rotating machinery experimental platform. The signals are sampled and divided into training, validation, and test sets and normalized. An imbalanced ratio Q is established to fit the data that is collected in actual environment (Eq. (11) below). Training sets are made into imbalanced training sets according to Q.

$$Q = \frac{N_{fault}}{N_{norm}} \quad (11)$$

A set of minority class data is generated in stage 2. In this way, a few samples are used for generation. In stage 3, the output data of stage 2 and original data are put into DSM

for screening. Data containing the most feature information of original data are found, and minority data are expanded with this set of data. In stage 4, including two sub stages: training stage and testing stage. The expanded balanced data sets and validation set are put into WDCNN for training in training stage. And the trained module is tested with test sets to obtain final diagnosis accuracy in testing stage. The intelligent fault algorithm for imbalanced data proposed in this paper is shown in the following algorithm 2. From algorithm 2, it can be seen that the specific derivation of the innovative implementation in Fig. 2. In training stage, the whole fault diagnosis accuracy should be acquired throughout four stages. But in testing stage, only original tested data are used in stage 4 and no more generated data generate by EGAN.


Algorithm 2 EGAN -DSM-WDCNN fault diagnosis for imbalanced data

Input: Imbalanced training sets X, $(x \in X)$, validation sets Y_1, test sets Y_2, gaussian noise n, random number $\varepsilon \sim U[0, 1]$, $\lambda = 10$, $j = 8000$

Output: fault diagnosis accuracy L

1: **for** $i = 1, 2...j$ **do**
2:     The loss value of generator $\tilde{g} = D(G(n_i))$
3:     $\hat{x} = \varepsilon x + (1 - \varepsilon)G(n_i)$
4:     The loss value of discriminator $\tilde{d} = D(x) - \tilde{g} - \lambda[(\left\| \nabla_{\hat{x}} D(\hat{x}) \right\|_2 - 1)^2]$
5:     Determine the difference between abs $(\tilde{g})$, abs $(\tilde{d})$ to target enhancement G or D.
6:     Generate data $G(n_i)$
7: **end for**
8: Screening data $G(n_i)_{best} = DSM(G(n_i))$
9: Mix data sets $X_{mix} = G(n_i)_{best} \bigcup X$
10: Input X_{mix}, Y_1, Y_2 into WDCNN to obtain accuracy L

Fig. 2 Fault diagnosis framework of EGAN-DSM-WDCNN
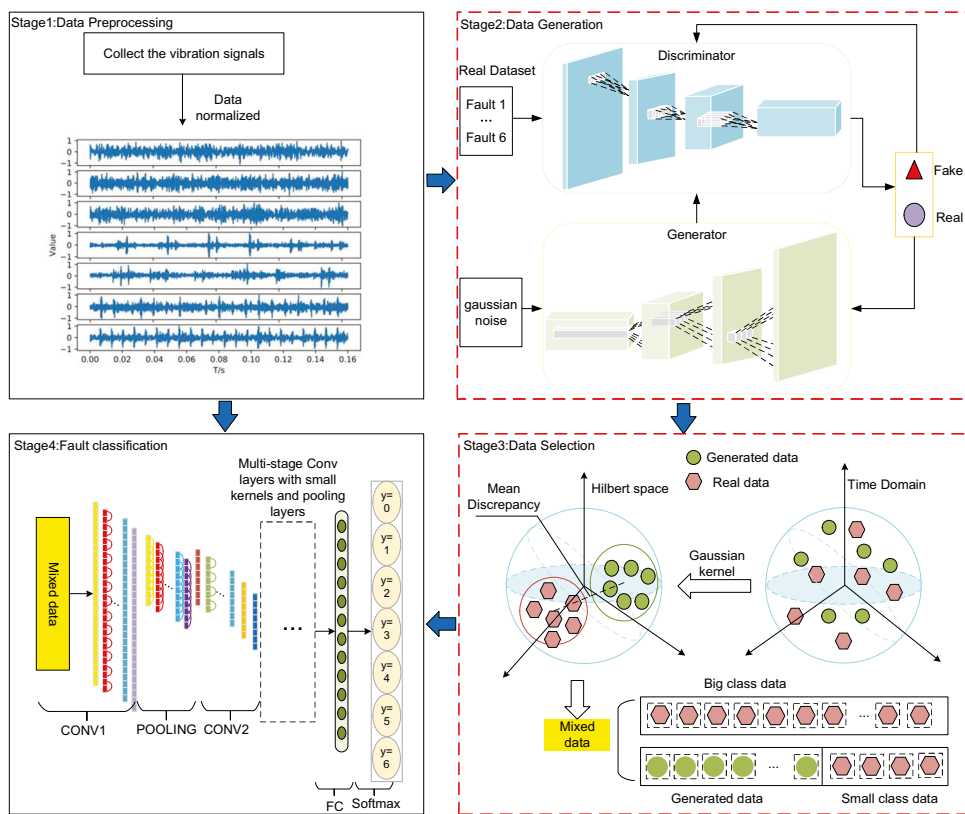

4 Experiments validation

4.1 Description of the data sets

In this section, two data sets are used validate the effectiveness of the proposed approach, including Henan University(HENU) bearing data sets and Case Western Reserve University(CWRU) bearing data sets.

4.1.1 HENU bearing data sets

The HENU bearing data sets are acquired by Information Fusion Lab, Henan University.The rotating machinery experimental platform is shown in Fig. 3. The whole experimental platform is mainly composed of a driving motor, speed sensor, motor controller, coupling, bearing base, acceleration sensor, load, bearings and so on. Bearings can be divided into

Fig. 3 Experimental platform of rotating machinery
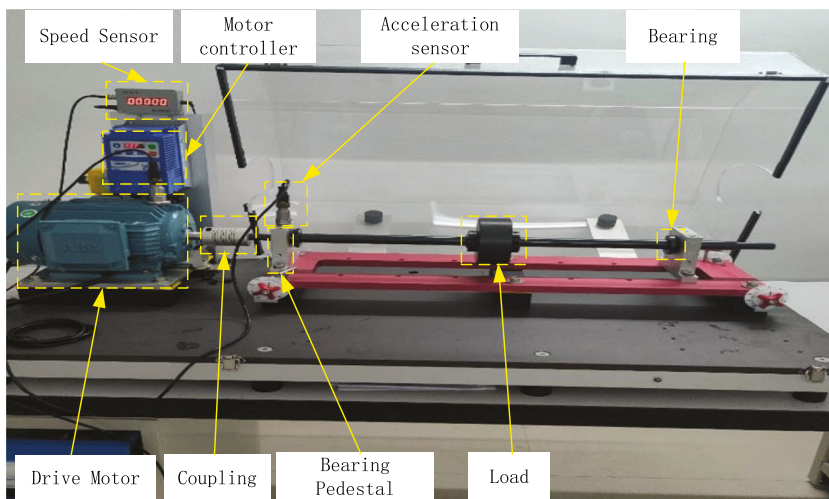
normal bearings and faulty bearings. Faulty bearings can be divided into 0.75 inch and 1 inch bearings according to the diameter of the bearing. Faulty bearings with same diameter can be divided into ball faulty bearings, outer faulty bearings, and inner faulty bearings. A motor controller is used to adjust the speed of driving motor. The coupling is connected to driving motor and rotating shaft. An acceleration sensor is used to collect vibration signals during operation of the equipment. Data acquisition unit transmits signal collected by acceleration sensor to computer.

Before experiment, the sampling frequency of the data acquisition unit is set as 12kHz by computer software; the motor speed is set as 2400 rpm by adjusting motor controller; and the acceleration sensor is installed on bearing base. During the experiment, two sizes of faulty bearings and three kinds of fault need to be measured. A total of seven different health states of bearing data including norm bearing data are collected. Specific data collection information is shown in Table 1.

4.1.2 CWRU bearing data sets

The CWRU bearing data sets are acquired by the Electrical Engineering Laboratory, Case Western Reserve University [2], and published on the Bearing Data Center Website. We chose the 100, 108, 121, 133, 172, 188, 200, 212, 225, 237 data. Among these ten classes bearing data, three different types of faults and three different fault diameters total of nine fault data and one norm data are included. Specific data collection information is shown in Table 2.

4.2 Performance metric

In this paper, the fault diagnosis of bearing imbalance data is a multiclass classification problem. To evaluate the performance of the model in the case of imbalance, we adopt the F-score and accuracy as the criterion. The F-score [44] is a comprehensive metric for evaluating the overall performance across all classes. The specific calculation formula is shown

Table 1 Description of HENU bearing data

Class label	Bearing diameter(inch)	Health state	Speed (rpm)	Load (kg)
Ball 0	0.75	ball fault	2400	5
Ball 1	1	ball fault	2400	5
Inner 0	0.75	inner fault	2400	5
Inner 1	1	inner fault	2400	5
Norm	0.75	norm	2400	5
Outer 0	0.75	outer fault	2400	5
Outer 1	1	outer fault	2400	5

Table 2 Description of CWRU bearing data

File number	Fault diameter(inch)	Health state	Speed (rpm)	Load (hp)
108	0.007	inner fault	1730	3
172	0.014	inner fault	1730	3
212	0.021	inner fault	1730	3
121	0.007	ball fault	1730	3
188	0.014	ball fault	1730	3
225	0.021	ball fault	1730	3
100	—	norm	1730	3
133	0.007	outer fault	1730	3
200	0.014	outer fault	1730	3
237	0.021	outer fault	1730	3

in Table 3, where tp, tn, fp, and fn refer to true positives, true negatives, false positives, and false negatives in a single class, respectively. Hereby, m is the number of classes to be classified.

4.3 Networks structure parameters selection

EGAN is based on the WGAN-GP model and enhances adversarial training of sub-neural networks. The parameters of generator and discriminator are shown in Table 4, which are determined by empirical values and a series of debugging experiments. The generator of EGAN consists of a fully connected layer and four 1-D deconvolution layers. The discriminator consists of four convolutional layers and one fully connected layer. The optimization function of loss function is Adam, whose learning rate, beta 1 and beta 2 are set as 0.0001, 0, and 0.9 respectively [41].

An environment of the whole experiment: CPU: Intel Xeno (R) W-2225, GPU: NVIDIA RTX A4000 and python 3.7.

4.4 Parameters selection experiments

4.4.1 Enhanced Threshold selection.

In EGAN, an Enhanced Threshold (ET) should be designed to judge the learning ability of two sub-neural networks. There-

Table 3 Performance metric and corresponding calculation expression

Performance Metric	Expression
Micro precision(MIP)	$\frac{\sum_{i=1}^{m} tp_i}{\sum_{i=1}^{m} (tp_i + fp_i)}$
Micro recall(MIR)	$\frac{\sum_{i=1}^{m} tp_i}{\sum_{i=1}^{m} (tp_i + fn_i)}$
F-score	$\frac{2(MIP*MIR)}{MIP+MIR}$

Table 4 EGAN parameters

	Layer name	Input	Output	Filter	AF	Stride
	Fc	[5,512]	[5,16,256]	—	relu	—
	Deconv1	[5,16,256]	[5,32,128]	[5,128,256]	relu	[1,2,1]
Generator	Deconv2	[5,32,128]	[5,128,64]	[5,64,128]	relu	[1,4,1]
	Deconv3	[5,128,64]	[5,512,32]	[5,32,64]	relu	[1,4,1]
	Deconv4	[5,512,32]	[5,2048,1]	[5,1,32]	tanh	[1,4,1]
	Conv1	[5,2048,1]	[5,512,32]	[5,1,32]	relu	[1,4,1]
	Conv2	[5,512,32]	[5,128,64]	[5,32,64]	relu	[1,4,1]
Discriminator	Conv3	[5,128,64]	[5,32,128]	[5,64,128]	relu	[1,4,1]
	Conv4	[5,32,128]	[5,16,256]	[5,128,256]	relu	[1,2,1]
	Flatten	[5,16,256]	[5,4096]	—	—	—
	Fc	[5,4096]	[5,1]	—	—	—

fore, experiments are needed to find an appropriate ET as a threshold for model enhancement training. According to the training experience of generative adversarial networks, ET is set to 1, 1.5, 2, 2.5, 3, and 3.5 respectively in HENU data sets.

It can be seen from Fig. 4 that the curve with ET of 2.5 is not inferior to other curves in minimum value. In addition, the curve is more stable which indirect the quality of generated data is better. At the same time, a comparison of curve 2.5 with curve 1 and curve 3.5 shows that whether ET is large or small will affect the quality of generated data from the perspective of stability. Besides, final diagnostic accuracy is used as an evaluation index for the selection of ET so as to increase the persuasiveness of ET. Diagnostic accuracy corresponding to different ET is shown in Fig. 5. In order to avoid the accident of experimental data, each accuracy in the figure is the average of ten results.

In Fig. 5, all accuracy is above 99% when Q is 1:10, indicating that ET has little effect on accuracy. However,

accuracy at an ET of 2.5 is significantly higher than the results of other ET when Q becomes smaller. The larger the Q, the more fault feature in training sets. Therefore, generated data improves the classification accuracy a little with a large Q. However, fault features in generated data will play a role due to the amount of fault features in original imbalanced data sets being small when Q is small. 2.5 is chosen as the value of ET during the experiments combined with the results obtained in Fig. 4.

4.4.2 Rated epoch selection.

Specifically, the total number of training epochs for EGAN is 8000, which would be costly if all the generated samples are recorded. So, it is necessary to choose a Rated Epoch (RE) to generate data. To find an appropriate RE, we compare the classification accuracy of generated data for different RE in HENU data sets. Average classification results are shown in Fig. 6.

Classification accuracy of different RE is basically above 99% when Q is 1:10. However, the accuracy gap becomes more and more obvious with decreasing of Q. And diagnostic
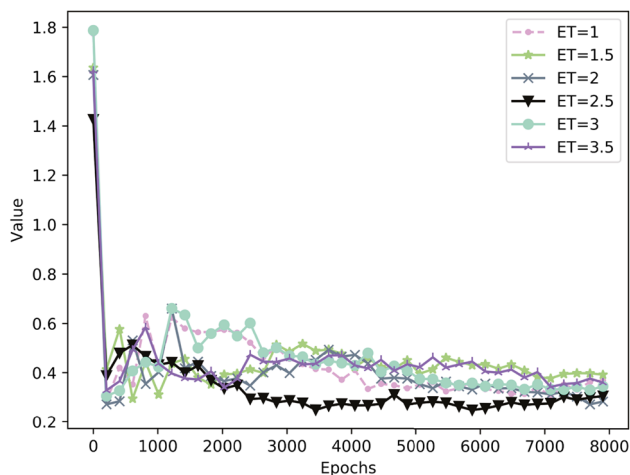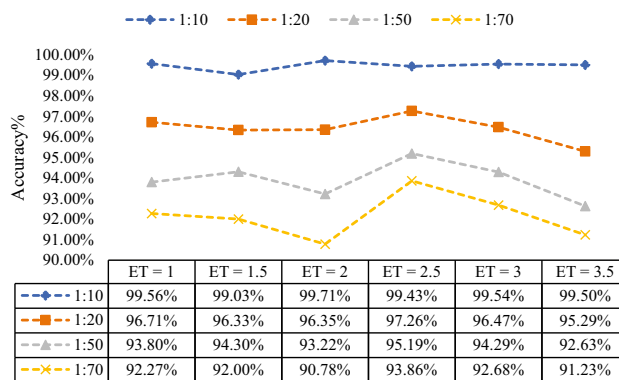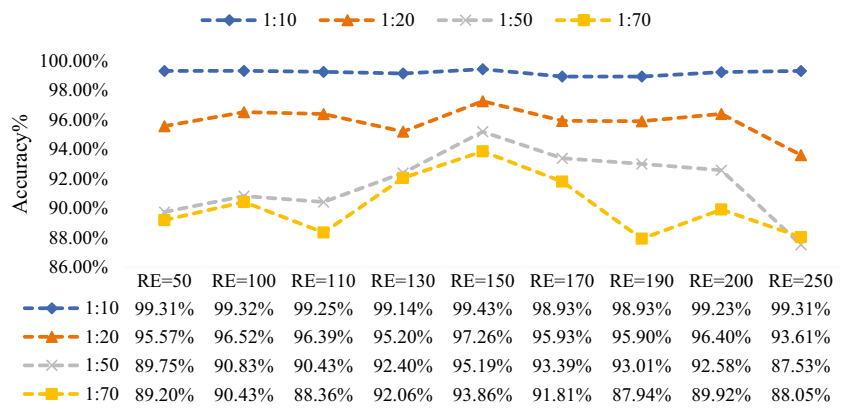

Fig. 4 MMD of generated data under different ET

	ET = 1	ET = 1.5	ET = 2	ET = 2.5	ET = 3	ET = 3.5
1:10	99.56%	99.03%	99.71%	99.43%	99.54%	99.50%
1:20	96.71%	96.33%	96.35%	97.26%	96.47%	95.29%
1:50	93.80%	94.30%	93.22%	95.19%	94.29%	92.63%
1:70	92.27%	92.00%	90.78%	93.86%	92.68%	91.23%

Fig. 5 Average accuracy and trend for different ET

Fig. 6 Average diagnostic accuracy and trends for different RE

	RE=50	RE=100	RE=110	RE=130	RE=150	RE=170	RE=190	RE=200	RE=250
1:10	99.31%	99.32%	99.25%	99.14%	99.43%	98.93%	98.93%	99.23%	99.31%
1:20	95.57%	96.52%	96.39%	95.20%	97.26%	95.93%	95.90%	96.40%	93.61%
1:50	89.75%	90.83%	90.43%	92.40%	95.19%	93.39%	93.01%	92.58%	87.53%
1:70	89.20%	90.43%	88.36%	92.06%	93.86%	91.81%	87.94%	89.92%	88.05%

accuracy reaches the maximum when RE is 150. The number of fault features contained in generated data of GAN is different for different accuracy in Fig. 6. That indirectly indicates the need for post-processing of generated data. Above all, 150 is chosen as the value of RE.

4.5 Ablation experiments

To demonstrate the validity of our method, this section shows ablation experiments at classification accuracy on HENU data sets. Q is selected as 1:20 for the purpose of presentation, and the results are shown in Table 5.

All accuracy rates in the table are calculated by WDCNN. Taking experiment 0 as a benchmark method, the multi-classification accuracy of bearing imbalanced data is 91.65%. As can be seen from Table 5, DSM is ablated in experiment 1 and the accuracy reaches 95.47%, which is 3.82% higher than that of benchmark method. In experiment 2, the enhanced processing module is ablated, and the accuracy reaches 95.2%, which is 3.55% higher than in experiment 0. The final experimental accuracy of experiment 3 reaches 97.26%, which is 5.61% higher than that of benchmark method. Therefore, a separate ablation experiment can verify that the proposed innovation improves the diagnostic accuracy of imbalanced data well.

4.6 Results and analysis

The whole experiment data consists of seven different states of bearing data, as shown in Table 6, where the example Q is set to 1:20. Q for whole experiments also includes 1:10, 1:50, and 1:70.

Sample lengths in Table 6 are sample points contained in one sample when original data is sampled. The training sets are sampled from original data under condition Q. And the input of original data is 25. Mixed sets are the number of training data sets after data enhancement. The test sets are sampled from original data and do not contain any generated data.

Table 5 Average accuracy of ablation experiments

experiment	Q	WGAN-GP	ET	DSM	Accuracy(%)
0	20	✓			91.65%
1	20	✓	✓		95.47%
2	20	✓		✓	95.20%
3	20	✓	✓	✓	97.26%

Table 6 Presentation of HENU data sets

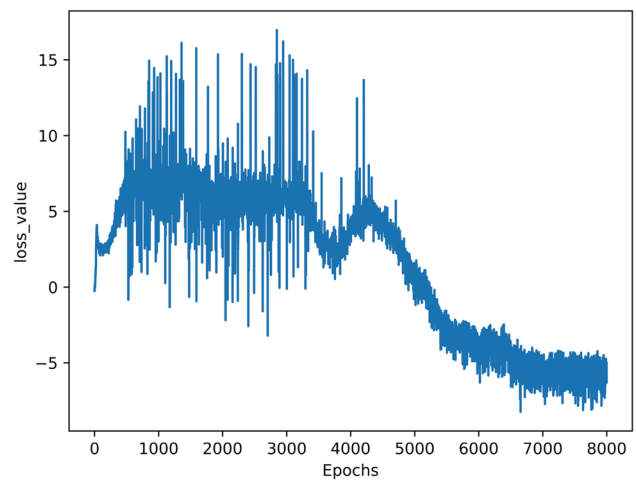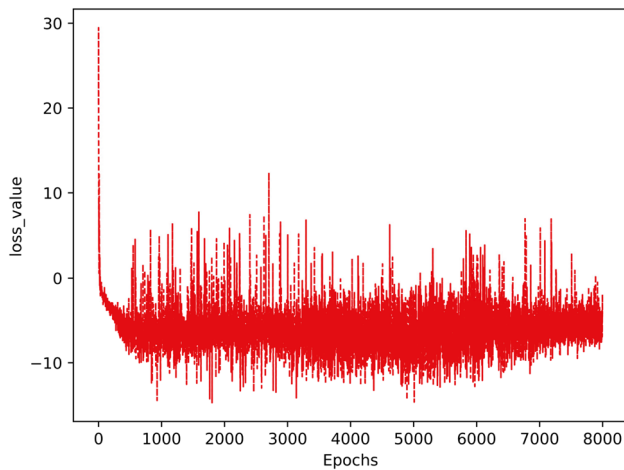| Label	Sample length	Training sets	Mixed sets	Test sets
Norm	2048	500	500	250
Ball 0	2048	25	500	250
Ball 1	2048	25	500	250
Inner 0	2048	25	500	250
Inner 1	2048	25	500	250
Outer 0	2048	25	500	250
Outer 1	2048	25	500	250

Fig. 7 The loss of Generator

Fig. 8 The loss of discriminator

4.6.1 EGAN results

Input data are 25 samples in Ball 0 of HENU data sets, and the EGAN model is trained to generate new samples for fault data. The model is completed after 8000 training sessions. Loss values of two sub-neural networks obtained at the end of each training session are shown in Fig. 7 and 8.

Generator and discriminator are optimized constantly and learn how to be better than the other as the number of training sessions increases. The loss value of generator decreases gradually, which means that the generator is capable of generating data increasingly after continuous learning. Meanwhile,

the distribution of generated data is closer and closer to real data. In Fig. 7, the loss fluctuates greatly from 0 to 5000 training times, but each large fluctuation is quickly smoothed out. This indirectly proves that the enhancement module proposed in this paper plays a role by enhancing the learning ability of weaker networks. That enables the whole model to resume adversarial training. Although the loss values of discriminator fluctuate sharply in Fig. 8, this indicates that the generator and discriminator train intensively against each other. In addition, the larger number of training sessions, the longer training time required.

Meanwhile, Fig. 9 shows the distribution of generated and real data in time domain. The red dashed line in each of these pictures represents generated fault data and the solid blue line indicates real fault data. The first row is the data comparison diagram of 0.75 inch faulty bearings. From left to right, the ball fault, inner fault, and outer fault are shown in sequence. The second row is 1 inch fault bearings.

The feature of real ball fault data is not obvious, which could make generation difficult. Between real data and generated data, the difference in data amplitude and deviation is the supplement of original fault features. The differences are that the original purpose of GAN is to expand training data features. Similarly, the supplementary performance of generated data on fault features is obvious because feature information of inner fault and outer fault are prominent. To sum up, generated data is similar to original data distribution from the perspective of overall trend. However, in terms of specific fault features, generated data contains richer feature information.
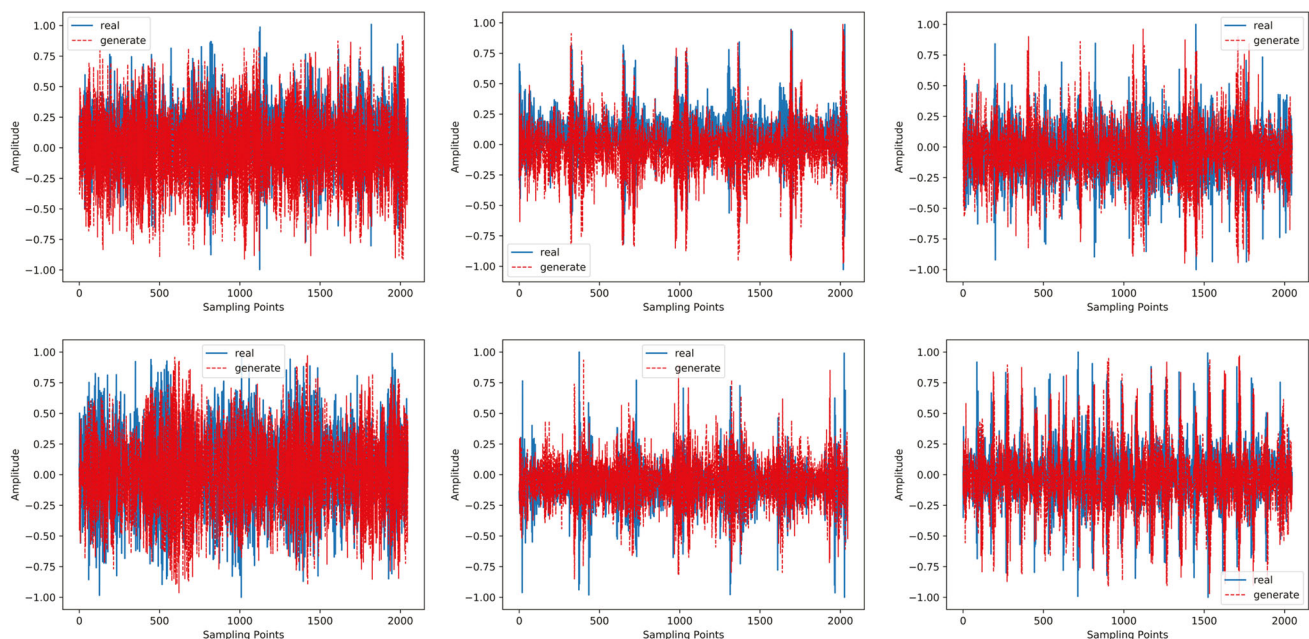

Fig. 9 Comparison of generated data with original data

Table 7 Average test accuracy and F-score for multiple Q

Q		1:10	1:20	1:50	1:70
un-processed	acc	98.29%	81.87%	26.36%	24.11%
	F-score	0.9839	0.8125	0.2064	0.1656
our-method	acc	99.43%	97.26%	95.19%	93.86%
	F-score	0.9979	0.9715	0.9634	0.9369

4.6.2 Fault diagnosis results presentation and analysis

In this section, the bearing data measured by the rotating machinery experimental platform is used to conduct the multi-classification experiments. The training data sets with different Q are used to verify the effectiveness of our method. The results are shown in Table 7. Un-processed means that there is no processing of imbalanced data sets.

Classification accuracy obtained by our method is improved by 1.14%, 15.39%, 68.83%, and 69.75% respectively at different Q. It can be seen that with the decreasing of Q, the diagnostic accuracy of WDCNN is significantly improved. Fault feature information in imbalanced data sets decreases with the decreasing of Q. When Q is small, the classifier cannot obtain enough feature information, and the diagnostic accuracy of the un-processed method is low. However, data enhanced by our method contains abundant fault feature information, which can help the classifier to distinguish test sets. Besides, the F-score is improved by 0.014, 0.159, 0.757, and 0.7713. The improvement of F-score confirms the effectiveness of the proposed method in dealing with an imbalanced dataset.

To demonstrate the validity of experiments, some parameters during training epochs are displayed in Fig. 10. Train loss and accuracy perform well in the whole epochs. But validation loss (val loss in Fig. 10) rises in early epochs and then
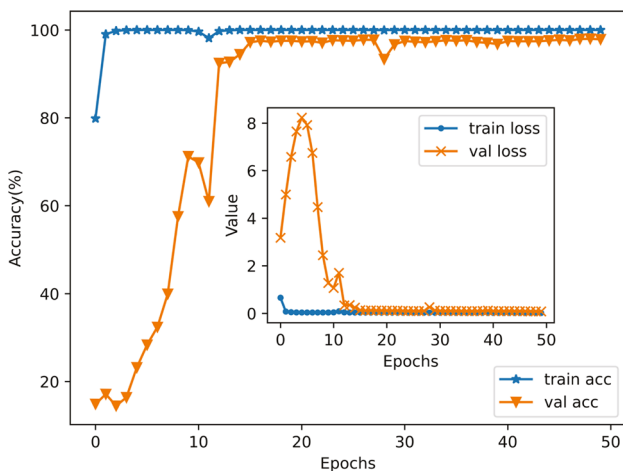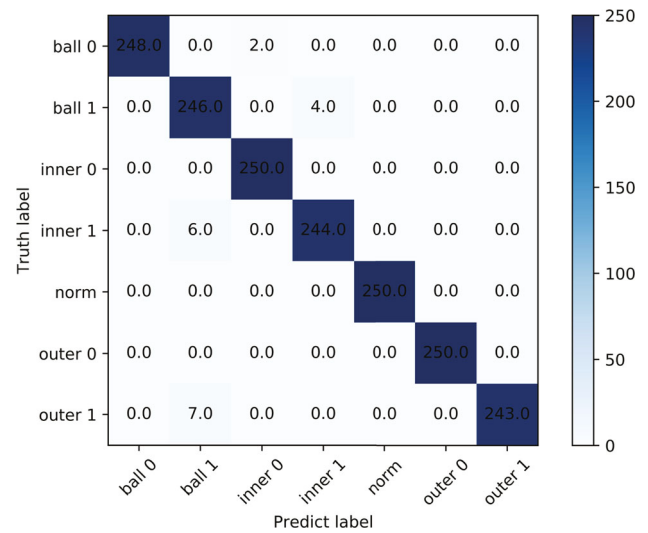

Fig. 11 Confusion matrix of our method

falls to a stable level. And validation accuracy (val acc in Fig. 10) rises gradually. In early epochs, some incorrectly predicted samples dominant val loss, resulting in val loss rises. But a large number of samples have been right predicted, which causes the accuracy curve rises. This phenomenon disappears as WDCNN training. As shown in Fig. 10, the val loss curve falls and the accuracy gets higher.

For the purpose of analyzing the classification accuracy of each type of data, Fig. 11 and 12 show the specific classification results for seven types of data when Q is 1:20 in the form of a confusion matrix.

Our method and un-processed method in Fig. 11 and 12 can be seen that the single-class accuracy of normal samples can reach 100%. However, for the diagnosis of other faults, the classification accuracy of our method is much
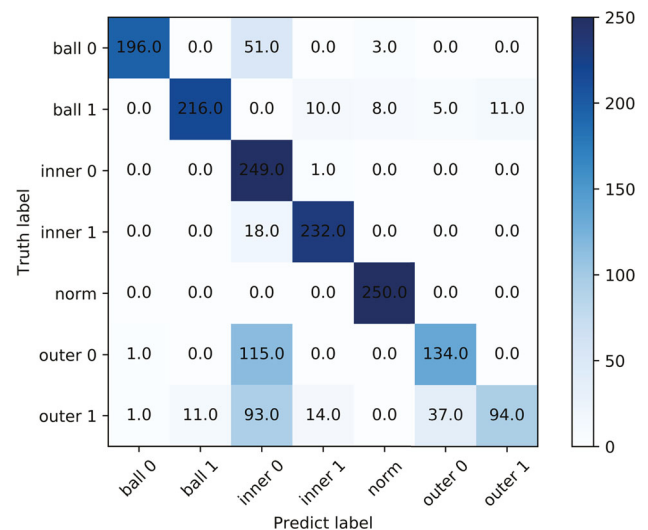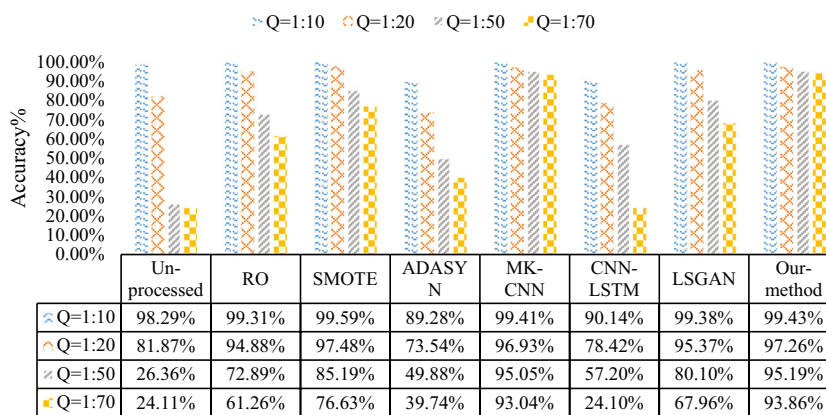

Fig. 10 Training results of WDCNN

Fig. 12 Confusion matrix of un-processed method

Fig. 13 Test accuracy of different methods at different Q

	Un-processed	RO	SMOTE	ADASYN	MK-CNN	CNN-LSTM	LSGAN	Our-method
Q=1:10	98.29%	99.31%	99.59%	89.28%	99.41%	90.14%	99.38%	99.43%
Q=1:20	81.87%	94.88%	97.48%	73.54%	96.93%	78.42%	95.37%	97.26%
Q=1:50	26.36%	72.89%	85.19%	49.88%	95.05%	57.20%	80.10%	95.19%
Q=1:70	24.11%	61.26%	76.63%	39.74%	93.04%	24.10%	67.96%	93.86%

higher than that of un-processed method. Specifically, the inner fault feature information in imbalanced data sets is less (it can be seen from Fig. 9 that the inner fault is similar to outer fault). Therefore, the misclassification of inner and outer fault in Fig. 12 is caused. However, the EGAN-DSM data augmentation method proposed in this paper can effectively supplement fault feature information. As can be seen in Fig. 11, the inner fault and outer fault are completely distinguished.

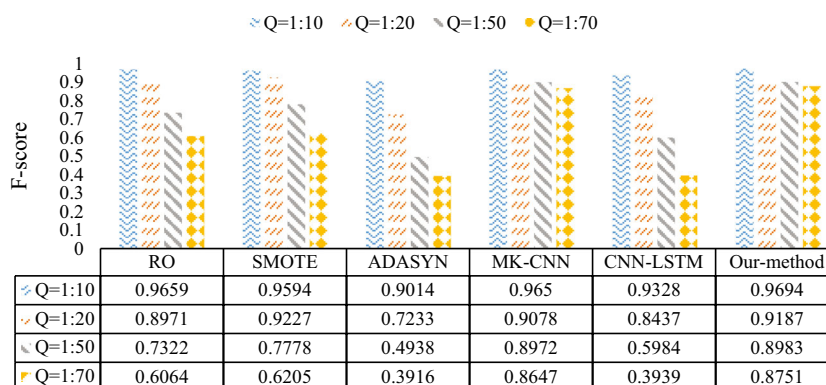### 4.6.3 Comparison results of different methods in different data sets

The imbalanced data reduces the accuracy of multiple classifications, so data augmentation is required to reduce the impact of imbalanced data sets. In this part, we discuss the accuracy of the proposed method and other data enhancement methods in multiple classifications on imbalanced data sets.

HENU data sets: As can be seen from Fig.13, Random Oversampling (RO) achieves the effect of data augmentation by randomly replicating minority class data so that the classification accuracy of this method is better when Q is high. When Q is small, the feature information of fault data

is still lacking after data enhancement, reducing the classification accuracy. ADASYN method is worse than other methods for classification accuracy. The data boundary of inner fault and outer fault is close. ADASYN adds different types of fault features when synthesizing data, which causes the final classification difficulty. CNN-LSTM [45] can efficiently extract temporal features, but imbalanced data characteristics impact it negatively. MK-CNN [46] focus on imbalanced feature through its multiscale kernel, which can get better classification accuracy. The method proposed in this paper does not widen the gap with other methods when Q is 1:10. Com-pared with SMOTE, the proposed method is even lower when Q is 1:10 and 1:20. The feature of generated data will be less critical when imbalanced training sets consist of an abundant feature. Our method is fit for highly imbalanced data sets which get close to actual industrial production. And the accuracy is much higher than other methods when Q is 1:50 and 1:70. The accuracy is higher than that of LSGAN, which indirectly indicates that the proposed method can generate and select higher quality data. Therefore, it can be shown that the EGAN-DSM method can effectively improve the performance of fault diagnosis under imbalanced conditions.

CWRU data sets: As seen in Fig.14, compared with RO, SMOTE, ADASYN, CNN-LSTM, and MK-CNN methods,

Fig. 14 F-score of different methods at different Q in CWRU data sets

	RO	SMOTE	ADASYN	MK-CNN	CNN-LSTM	Our-method
Q=1:10	0.9659	0.9594	0.9014	0.965	0.9328	0.9694
Q=1:20	0.8971	0.9227	0.7233	0.9078	0.8437	0.9187
Q=1:50	0.7322	0.7778	0.4938	0.8972	0.5984	0.8983
Q=1:70	0.6064	0.6205	0.3916	0.8647	0.3939	0.8751

our method achieved the best F-score in CWRU data sets. When the Q is large, the F-score of the pro-posed method and other methods are not far apart. However, as the Q decreases, the difference between the F-score of the proposed method and other methods becomes larger and larger. This can indicate directly that our method can solve imbalance bearing data effectively.

5 Conclusion

Imbalanced data limits the performance of intelligent diagnostic methods, resulting in lower diagnostic accuracy. In order to reduce the influence of bearing imbalance data, we propose an EGAN-DSM data enhancement method that achieves excellent performance on fault diagnosis. EGAN adds an enhancement module to WGAN-GP which can effectively avoid antagonistic suppression during training and improve the quality of generated data. DSM can effectively screen generated data containing the most original fault features in the training process and enrich the feature information in the mixed sets. Compared with other methods (MK-CNN, SMOTE, ADASYN, RO, CNN-LSTM, LSGAN), EGAN-DSM performs better in different fault diagnosis experiments. Although our approach may not outperform other methods at larger Q, the difference is minimal. In addition, our method is far superior to other methods when Q is small. In future work, the influence of data enhancement methods on fault diagnosis accuracy under a strong noise environment will be investigated, including new GAN construction and efficient classification networks construction.

Author Contributions Yandong Hou: Methodology, Software, Investigation. Jiulong Ma: Methodology, Conceptualization, Supervision, Writing original draft, Software, Investigation. Zhengquan Chen: Validation, Formal analysis, Writing-review and editing. Jinjin Wang: Graphing, Writing-review and editing. Tianzhi Li: Validation, Writing-review and editing.

Data Availability The HENU datasets generated during the current study are available from the corresponding author on reasonable request. The CWRU datasets can be accessed with https://engineering.case.edu/bearingdatacenter/download-data-file

Declarations

Conflicts of interest The authors have no financial or proprietary interests in any material discussed in this article.

Ethical standard The data used in this paper has no potential conflicts of interest. And the data used have obtained permission to use.

References

1. Deng W, Li Z, Li X et al (2022) Compound fault diagnosis using optimized mckd and sparse representation for rolling bearings. IEEE Trans Instrum Meas. 71:1–9. https://doi.org/10.1109/TIM.2022.3159005
2. Ma J, Shang J, Zhao X et al (2022) Bayes-dcgru with bayesian optimization for rolling bearing fault diagnosis. Appl Intell. 52:11,172–11,183. https://doi.org/10.1007/s10489-021-02924-z
3. Shao H, Xia M, Wan J et al (2022) Modified stacked autoencoder using adaptive morlet wavelet for intelligent fault diagnosis of rotating machinery. IEEE/ASME Transactions on Mechatronics 27(1):24–33. https://doi.org/10.1109/TMECH.2021.3058061
4. Wang Z, Yang J, Guo Y (2022) Unknown fault feature extraction of rolling bearings under variable speed conditions based on statistical complexity measures. Mechanical Systems and Signal Processing 172(108):964. https://doi.org/10.1016/j.ymssp.2022.108964
5. Bendjama H (2022) Bearing fault diagnosis based on optimal morlet wavelet filter and teager-kaiser energy operator. Journal of the Brazilian Society of Mechanical Sciences and Engineering 44(9):1–23. https://doi.org/10.1007/s40430-022-03688-4
6. Hoang DT, Kang HJ (2019) A survey on deep learning based bearing fault diagnosis. Neurocomputing 335:327–335. https://doi.org/10.1016/j.neucom.2018.06.078
7. Liu H, Nie H, Zhang Z et al (2020) Anisotropic angle distribution learning for head pose estimation and attention understanding in human-computer interaction. Neurocomputing 433:310–322. https://doi.org/10.1016/j.neucom.2020.09.068
8. Li Z, Liu H, Zhang Z et al (2022) Learning knowledge graph embedding with heterogeneous relation attention networks. IEEE Transactions on Neural Networks and Learning Systems 33(8):3961–3973. https://doi.org/10.1109/TNNLS.2021.3055147
9. Liu T, Wang J, Yang B et al (2021) Ngdnet: Nonuniform gaussian-label distribution learning for infrared head pose estimation and on-task behavior understanding in the classroom. Neurocomputing 436(4):210–220. https://doi.org/10.1016/j.neucom.2020.12.090
10. Li X, Li T, Li S et al (2023) Learning fusion feature representation for garbage image classification model in human-robot interaction. Infrared Physics and Technology 128(104):457. https://doi.org/10.1016/j.infrared.2022.104457
11. Wang X, Mao D, Li X (2021) Bearing fault diagnosis based on vibro-acoustic data fusion and 1d-cnn network. Measurement 173(108):518. https://doi.org/10.1016/j.measurement.2020.108518
12. Xing S, Lei Y, Wang S et al (2020) Distribution-invariant deep belief network for intelligent fault diagnosis of machines under new working conditions. IEEE Trans Ind Electron. 68(3):2617–2625. https://doi.org/10.1109/TIE.2020.2972461
13. Cui M, Wang Y, Lin X et al (2021) Fault diagnosis of rolling bearings based on an improved stack autoencoder and support vector machine. IEEE Sensors J. 21(4):4927–4937. https://doi.org/10.1109/JSEN.2020.3030910
14. Liu H, Fang S, Zhang Z et al (2022) Mfdnet: Collaborative poses perception and matrix fisher distribution for head pose estimation. IEEE Transactions on Multimedia 24:2449–2460. https://doi.org/10.1109/TMM.2021.3081873
15. Liu H, Liu T, Zhang Z et al (2022) Arhpe: Asymmetric relation-aware representation learning for head pose estimation in industrial human-computer interaction. IEEE Transactions on Industrial Informatics 18(10):7107–7117. https://doi.org/10.1109/TII.2022.3143605
16. Liu H, Zheng C, Li D et al (2022) Edmf: Efficient deep matrix factorization with review feature learning for industrial recommender system. IEEE Trans Ind Inform. 18(7):4361–4371. https://doi.org/10.1109/TII.2021.3128240

17. Liu H, Zhang C, Deng Y et al (2023) Transifc: Invariant cues-aware feature concentration learning for efficient fine-grained bird image classification. IEEE Transactions on Multimedia pp 1–14. https://doi.org/10.1109/TMM.2023.3238548

18. Liu H, Liu T, Chen Y, et al (2022) Ehpe: Skeleton cues-based gaussian coordinate encoding for efficient human pose estimation. IEEE Transactions on Multimedia, pp 1–12. https://doi.org/10.1109/TMM.2022.3197364

19. Huang K, Wu S, Li F et al (2022) Fault diagnosis of hydraulic systems based on deep learning model with multirate data samples. IEEE Trans Neural Netw Learn Syst. 33(11):6789–6801. https://doi.org/10.1109/TNNLS.2021.3083401

20. Zareapoor M, Shamsolmoali P, Yang J (2021) Oversampling adversarial network for class-imbalanced fault diagnosis. Mechanical Systems and Signal Processing 149(107):175. https://doi.org/10.1016/j.ymssp.2020.107175

21. Yan K, Su J, Huang J et al (2022) Chiller fault diagnosis based on vae-enabled generative adversarial networks. IEEE Trans Autom Sci Eng. 19(1):387–395. https://doi.org/10.1109/TASE.2020.3035620

22. Zang Y, Zhang Z, Zhao X et al (2022) Bearing fault diagnosis method based on vae gan and flcnn unbalanced samples. J Vib Shock. 41:199–209. https://doi.org/10.13465/j.cnki.jvs.2022.09.026

23. Zhang T, Chen J, Li F et al (2022) Intelligent fault diagnosis of machines with small and imbalanced data: A state-of-the-art review and possible extensions. ISA Transactions 119:152–171. https://doi.org/10.1016/j.isatra.2021.02.042

24. Chawla NV, Bowyer KW, Hall LO et al (2002) Smote: synthetic minority oversampling technique. J artif intell res. 16:321–357. https://doi.org/10.1613/jair.953

25. Han H, Wang WY, Mao BH (2005) Borderline-smote: a new over-sampling method in imbalanced data sets learning. In: Int conf intell comput. Springer, pp 878–887,https://doi.org/10.1007/11538059_91

26. He H, Bai Y, Garcia EA, et al (2008) Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp 1322–1328, https://doi.org/10.1109/IJCNN.2008.4633969

27. Li Z, Zheng T, Wang Y et al (2021) A novel method for imbalanced fault diagnosis of rotating machinery based on generative adversarial networks. IEEE Trans Instrum Meas. 70:1–17. https://doi.org/10.1109/TIM.2020.3009343

28. Pan T, Chen J, Zhang T et al (2022) Generative adversarial network in mechanical fault diagnosis under small sample: A systematic review on applications and future perspectives. ISA Transactions. 128:1–10. https://doi.org/10.1016/j.isatra.2021.11.040

29. Zhu J, Jiang M, Liu Z (2022) Fault detection and diagnosis in industrial processes with variational autoencoder: A comprehensive study. Sensors 22(1):227. https://doi.org/10.3390/s22010227

30. Goodfellow I, Pouget-Abadie J, Mirza M et al (2020) Generative adversarial networks. Communications of the ACM. 63(11):139–144. https://doi.org/10.1145/3422622

31. Zhao D, Liu S, Gu D et al (2019) Enhanced data-driven fault diagnosis for machines with small and unbalanced data based on variational auto-encoder. Meas Sci Technol. 31(3):035,004. https://doi.org/10.1088/1361-6501/ab55f8

32. Li M, Zou D, Luo S et al (2022) A new generative adversarial network based imbalanced fault diagnosis method. Measurement 194(111):045. https://doi.org/10.1016/j.measurement.2022.111045

33. Liu J, Zhang C, Jiang X (2022) Imbalanced fault diagnosis of rolling bearing using improved msr-gan and feature enhancement-driven capsnet. Mech Syst Signal Proc. 168(108):664. https://doi.org/10.1016/j.ymssp.2021.108664

34. Shao S, Wang P, Yan R (2019) Generative adversarial networks for data augmentation in machine fault diagnosis. Computers in Industry 106:85–93. https://doi.org/10.1016/j.compind.2019.01.001

35. Zhang W, Li X, Jia XD et al (2020) Machinery fault diagnosis with imbalanced data using deep generative adversarial networks. Measurement 152(107):377. https://doi.org/10.1016/j.measurement.2019.107377

36. Hao D, Gao X, Qi W (2022) Data augmentation method based on improved generative adversarial network for the sucker rod pump system. Int J Control, Autom Syst. 20(11):3718–3730. https://doi.org/10.1007/s12555-021-0691-y

37. Yang G, Zhong Y, Yang L et al (2021) Fault diagnosis of harmonic drive with imbalanced data using generative adversarial network. IEEE Trans Instrum Meas. 70:1–11. https://doi.org/10.1109/TIM.2021.3089240

38. Li X, Zhang W, Ding Q et al (2019) Multilayer domain adaptation method for rolling bearing fault diagnosis. Signal Processing 157:180–197. https://doi.org/10.1016/j.sigpro.2018.12.005

39. Gao X, Deng F, Yue X (2020) Data augmentation in fault diagnosis based on the wasserstein generative adversarial network with gradient penalty. Neurocomputing 396:487–494. https://doi.org/10.1016/j.neucom.2018.10.109

40. Arjovsky M, Chintala S, Bottou L (2017) Wasserstein generative adversarial networks. In: International conference on machine learning, PMLR, pp 214–223, https://dl.acm.org/doi/abs/10.5555/3305381.3305404

41. Gulrajani I, Ahmed F, Arjovsky M et al (2017) Improved training of wasserstein gans. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc, p 5769–5779, https://dl.acm.org/doi/10.5555/3295222.3295327

42. Zhang W, Peng G, Li C et al (2017) A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors 17(2):425. https://doi.org/10.3390/s17020425

43. Gretton A, Borgwardt KM, Rasch MJ et al (2012) A kernel two-sample test. The J Mach Lear Res. 13(1):723–773. https://dl.acm.org/doi/10.5555/2188385.2188410

44. He Q, Pang Y, Jiang G et al (2021) A spatiotemporal multiscale neural network approach for wind turbine fault diagnosis with imbalanced scada data. IEEE Trans Ind Inform. 17(10):6875–6884. https://doi.org/10.1109/TII.2020.3041114

45. Zhi Z, Liu L, Liu D et al (2022) Fault detection of the harmonic reducer based on cnnlstm with a novel denoising algorithm. IEEE Sensors Journal 22(3):2572–2581. https://doi.org/10.1109/JSEN.2021.3137992

46. Liu R, Wang F, Yang B et al (2020) Multiscale kernel based residual convolutional neural network for motor fault diagnosis under non-stationary conditions. IEEE Trans Ind Inform. 16(6):3797–3806. https://doi.org/10.1109/TII.2019.2941868

Yandong Hou received his Ph.D. degree in power electronics and power transmission from Shanghai Maritime University, Shanghai, China, in 2010. Currently, he is a Professor in Henan University, Zhengzhou, Henan. His research interests include fault diagnosis and Safety prediction in complex systems.
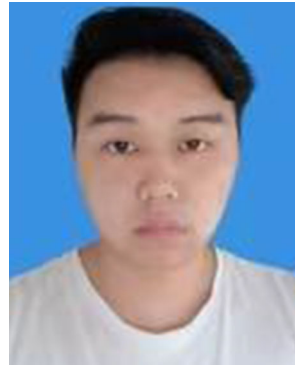
Tianzhi Li received his B.E. degree in automation from Technology & Media University of Henan Kaifeng, China, in 2022. He is currently pursuing BM degree in School of Artificial Intelligence, Henan University, China. Currently, his research interests include fault diagnosis, gan and imbalanced datasets.

Jiulong Ma received his B.E. degree in automation from Henan University, China, in 2021. He is currently pursuing BM degree in School of Artificial Intelligence, Henan University, China. Currently, his research interests include fault diagnosis, gan and imbalanced datasets.

Zhengquan Chen received his M.E. degree in control theory and control engineering from Henan University, China, in 2020. He is currently a Ph.D. student in School of Computer and Information Engineering, Henan University, China. Currently, his research interests include fault diagnosis, iterate learning control and robust.

Jinjin Wang received her B.E. degree in Communication Engineering from Shanghai Polytechnic University, China, in 2020. She is currently pursuing BM degree in School of Artificial Intelligence, Henan University, China. Currently, her research interests include fault diagnosis, and deep learning.