Research on Fault Diagnosis based on Improved Generative Adversarial Network under Small Samples

Li Dongping, Yang Yingchun*, Shen Shikai, He Jun, Shen Haoru, Yue Qiang, Hong Sunyan and Deng Fei

***Abstract*—Fault diagnosis based on deep learning has become a research hotspot. However, the classification accuracy may be low if the amount of training data is insufficient. To solve this problem, a new GAN generator is designed in this paper. The original faulty sample is replaced with the unbalanced faulty feature, which is used as the generator input. Instead of taking the data distribution consistency as the loss term, the faulty feature and diagnosis error are added to the training loss function. Bayesian optimization strategy is used to adjust the hyperparameters of the discriminator. The self-attention mechanism is introduced to increase the calculation of long and local correlation. Experiments show that the proposed method maintains high recognition accuracy for small samples.**

***Index Terms*—deep learning, small samples, fault diagnosis, pattern recognition**

I. Introduction

IN the process of industrial production, the failure frequency of some systems is relatively high. The faulty cause and type are diverse. Improving fault diagnosis ability has theoretical and practical significance to ensure the safety of industrial production.

As a common fault diagnosis method, pattern recognition algorithms mainly include extreme learning machine (ELM) [1], support vector machine (SVM) [2], convolutional neural network (CNN) [3], etc. ELM has fast learning speed and realizes rapid fault diagnosis. But the stability of ELM is relatively weak [4]. SVM can efficiently solve nonlinear decision-making problem with high-dimension. But it is difficult to select kernel parameters. The diagnostic accuracy is significantly affected by fault samples [5]. CNN, as a deep learning method, conducts network based on the original faulty samples. CNN establishes the mapping relationship between faulty samples and categories. The observed samples are trained for fault judgment. However, it is difficult to effectively use CNN with a small amount of faulty samples [6]. The existing research of fault diagnosis has achieved good performance in the experimental environment. Although deep learning such as CNN has made some achievements for fault diagnosis, there are still many challenges. Firstly, a large amount of high-quality data is the guarantee of good performance. However, the number of faulty data is very small, especially for catastrophic or accidental fault. And the number of normal data is very large [7]. Therefore, data imbalance is a common problem for fault diagnosis. If the number of faulty data is small, the fault diagnosis performance based on deep learning may be poor [8], resulting in limited recognition accuracy. Fault diagnosis with small samples has become an important research field in data-driven fault diagnosis.

There are mainly two fault diagnosis methods under small samples. The first method focuses on the cost sensitive learning. This method makes the network structure more sensitive to the classes with a few samples. However, there are two difficulties: (1) The cost of error classification cannot be determined. (2) Evaluating the performance of costing sensitive classifiers is difficult. The second method focuses on using some data preprocessing techniques, such as over-sampling and down-sampling, to reduce the imbalance between classes [9-13]. A simple over-sampling method is to copy several samples to increase the sample number of small classes. However, it is prone to the problem of over-fitting. For down-sampling preprocessing, it reduces the sample number of big classes, which may inevitably cause the problem of information loss.

Goodflow et al. proposed generative adversarial network (GAN). GAN uses the confrontation between the generator and discriminator to continuously optimize the network [14]. GAN has the superior ability to learn feature. Initially, GAN is mainly used in data and image generation. To increase sample number of other data type, GAN derivative models have been put forward. Semi supervised GAN is proposed in recent years [15-18]. The structure of semi supervised GAN

Manuscript received November 8, 2021; revised October 22, 2022. This work was supported in part by the General Project of Universities in Yunnan 201801CF00022, named “Research on Feature Learning Methods Based on Deep Computing in Big Data Environment” and the National Natural Science Foundation of China 62066023, named “Research on Automatic Annotation Technology of Yi Speech Data Based on Poverty Alleviation Log”.

Li Dongping is an associate professor of Institute of Information Engineering, Kunming University, Kunming, China (email: ynkm1971@sina.com)

Yang Yingchun, the corresponding author, is a senior engineer of Yunnan Branch of China Telecom Co., Ltd., Kunming, China. (e-mail: gddx89@21cn.com).

Shen Shikai is a professor of Institute of Information Engineering, Kunming University, Kunming, China. (e-mail: liamlychee@163.com).

He Jun is a professor of Institute of Information Engineering, Kunming University, Kunming, China. (e-mail: 2021kmxy@sina.com).

Shen Haoru is a lecturer of Institute of Information Engineering, Kunming University, Kunming, China. (lu20190816@126.com).

Yue Qiang is an associate professor of Institute of Information Engineering, Kunming University, Kunming, China. (e-mail: yueqiang2021@126.com).

Hong Sunyan is a lecturer of Institute of Information Engineering, Kunming University, Kunming, China. (e-mail: kmuhongsunyan@126.com).

Deng Fei is a senior engineer of Institute of Information Engineering, Kunming University, Kunming, China. (e-mail: feigeoffice@163.com).

is similar to the traditional GAN. The generator structure remains unchanged. The discriminator adds an output layer to achieve a multi classification function. The semi supervised GAN has been gradually applied to the fields of speech recognition, image recognition and sample classification, etc.

Although the above research has achieved some achievements, there are still some deficiencies in the training process: (1) The model may produce meaningless noise due to the small size of training sample. GAN can not accurately learn the distribution of real data. (2) GAN has the problem of gradient disappearance, which may lead to training failure of model. If the sample quality is poor, fault diagnosis with the generated samples is invalid. (3) The GAN hyperparameter optimization is cumbersome, which affects the discriminator performance.

To solve the above problems, a new GAN generator is designed to generate samples. Faulty characteristic extracted by the convolutional neural network is used as the generator input. The faulty characteristics and diagnostic errors are added to the training loss function of the generator. Data distribution consistency is taken as the loss item. Bayesian optimization strategy is used to adjust the hyperparameters of the discriminator. The subsequent sections are organized as follows: The GAN is briefly introduced in Section 2. The improved GAN is proposed in Section 3, which includes generator design, loss function, Bayesian optimization and self-attention mechanism. The fault diagnosis flow based on the proposed method is given in Section 4. The proposed method is verified through experiment in Section 5. The conclusion is conducted in Section 6.

II. Generative Adversarial Network

Generative adversarial network (GAN) model is usually composed of the generator and discriminator. GAN is proposed from the idea of zero-sum game. The generator continuously learns the prior distribution of true samples in the training set. Then the generator generates false samples with the same distribution. Discriminator judges whether the generated samples are true or false. The generator and discriminator are trained at the same time to finally achieve Nash equilibrium. The basic structure of GAN is shown in Figure 1. The noise is imported into the generator. The data with the same distribution of true sample is generated by the generator. True sample and generated data are two inputs for discriminator. The discriminator is equivalent to a two-class classifier. Through training the neural network, the probability that the input data belongs to the real sample is given. The GAN training process is the game between the generator and discriminator. The objective function is as the following.

$$\min_G \max_D L(D,G) = E_{x\sim P_{\text{real}}}[\log D(x)] + E_{y\sim P_{gen}}[\log(1-D(G(y)))] \quad (1)$$

where x is the real sample data. y is the random noise. P_{real} is the generator distribution on data x. P_{gen} is the discriminator distribution on data y. $D(x)$ is the probability of judging the input data as a real sample. $G(y)$ is the generated sample with random noise y. $D(x)$ is also written as $D(x;\theta_d)$, indicating the probability that x comes from real data rather than the generator. $G(y)$ is also written as $G(y;\theta_g)$, where $G()$ is the differential function represented by a multilayer perceptron with parameter θ_g.

To distinguish the real samples, the discriminator keeps $D(x)$ close to 1 while $D(G(y))$ close to 0 during the training process. The $\max_D$ is computed to increase $L(D,G)$. To produce more real sample, the generator keeps $D(G(y))$ close to 1. The $\min_G$ is calculated to diminish $L(D,G)$. Finally, multiple cross training is performed between the generator and the discriminator. And samples with the same distribution of the real samples are generated by GAN.

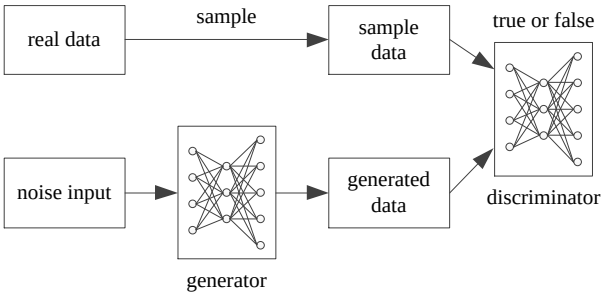


Fig.1 GAN structure

III. The Improved Generative Adversarial Network (IMGAN)

A. Design of Generator and Loss Function

1. Generator Design

To use the unbalanced samples, convolutional neural network is utilized to extract the sample features. The loss function is added in the generator. Then the generated features are decoded to produce samples. The specific process is as follows.

Firstly, random number y is input into the convolutional neural network to generate x_{fake}.

$$s_y = f_{\alpha_{G1}}(y) = \lambda_1(q_{G1}y + c_{G1}) \quad (2)$$

$$f_{fake} = f_{\alpha_{G2}}(s_y) = \lambda_1(q_{G2}h_y + c_{G2}) \quad (3)$$

where $\alpha_{G1} = \{q_{G1}, c_{G1}\}$ is used to connect the input and hidden layer. $\alpha_{G2} = \{q_{G2}, c_{G2}\}$ is used to connect the hidden and output layer of convolutional neural network (CNN) respectively. q_{G1} and q_{G2} are the weight matrix. c_{G1} and c_{G2} are the deviation. λ_1 is the activation function of hyperbolic tangent.

To obtain the generated sample x_{fake}, f_{fake} is decoded. The automatic encoder is used to encode f_{fake}. The unbalanced sample x_{real} is used to train CNN and the encoder. The specific process can be expressed as the following formula.

$$f_{real} = f_{\alpha_{CNN}}(x_{real}) = \lambda_1(q_{CNN}x_{real} + c_{CNN}) \quad (4)$$

$$\hat{x}_{real} = g_{\alpha_{AE}}(x_{real}) = \lambda_1(q_{AE}^{\mathrm{T}}x_{real} + c_{AE}) \quad (5)$$

where f_{real} is the feature vector of x_{real}. $\alpha_{CNN} = \{q_{CNN}, c_{CNN}\}$ is the parameter of the CNN feature extraction network. $\alpha_{AE} = \{q_{AE}, c_{AE}\}$ is the corresponding decoder network parameter. The decoding process on f_{fake} can be expressed as the following.

$$x_{fake} = g_{AE}(f_{fake}) = \lambda_1(q_{AE}^{T} f_{fake} + c_{AE}) \quad (6)$$

There are three neural network layers in the generator. The randomly generated noise is taken into the generator. Faulty samples are generated after deconvolution. Each neural network layer includes transposed convolution, batch normalization, and activation functions. The input dimension is changed by setting multiple 4 × 4 deconvolution kernels. The first transposed convolution expands the input noise dimension from 100 to 256. Then the second and third transposed convolution layers progressively reduce the dimension by half. The standardization process is proceeded for each transposed convolution. Bounded function is helpful to cover the training distribution space. As a result, tanh is used as the activation function in the last layer. ReLU is used as the activation function in the other three transposed convolution layers.

2. Loss Function Design

For the traditional GAN, the generator parameters are optimized by minimizing the loss function of cross entropy, which ensures the distribution consistency of the original sample x_{real} and the generated sample x_{fake}.

$$Q_g = -\frac{1}{N}\sum_{i=1}^{N} \log c_{fake} \quad (7)$$

To generate more qualified samples, the eigenvector f_{real} is used for the generator training. By minimizing the new loss function defined in formula (8), the generator parameters can be optimized as the following.

$$Q_G = Q_g + Q_f + Q_e \quad (8)$$

$$Q_f = \frac{1}{M}\left\| f_{real} - f_{fake} \right\|^2 \quad (9)$$

$$Q_e = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} -(z_{ik})\log(\hat{z}_{ik}) - (1-z_{ik})\log(1-\hat{z}_{ik}) \quad (10)$$

where Q_g can learn the distribution of the original samples. Q_f makes the characteristics of the generated sample close to the original sample. Q_e makes the generated samples meet the ultimate purpose. Adding Q_f and Q_e to the loss function is conducive to fault diagnosis.

3. Discriminator Design

The features of the input faulty samples are extracted through three convolution layers. Different 4 × 4 convolution kernels are used by all the convolution layers. The pooling layer is not utilized. Thus, a large search range is obtained by the convolution layer. The input dimension of each discriminator layer is opposite to the input dimension of each generator layer. The Leaky ReLU is used as the activation function for all the layers. The sigmoid function is used to convert the vector into a scalar output (0, 1), which represents the probability that the sample is real.

B. Bayesian Optimization

Usually, GAN hyperparameters include the number of hidden unit, the size of convolution kernel, pooling factor, etc. Selecting the appropriate hyperparameters is a cumbersome process. The common methods include empirical setting, grid search and Bayesian optimization. The empirical setting method is sensitive to human experience. The grid search method takes a long time to train the model. And it may be encountered by the problem of gradient explosion. The Bayesian optimization algorithm learns the objective function according to the characteristics. It generally includes two parts: the probability model and the acquisition function.

1. Probability Model

As a priori function, the probability model represents the unknown objective function under evaluation. After continuous and iterative calculation, the parameters are updated to modify the probability model. The Gaussian process is selected due to its high flexibility and strong expansibility. The expression is shown in (11).

$$f(x) \sim P(n(x), v(x, x')) \quad (11)$$

where x is the sample of the training set $X=\{x_1, x_2, \ldots, x_m\}$. $v(x, x')$ is the covariance function. $f(x)$ is the evaluation of the unknown function corresponding to the sample. $n(x)$ is the mean function of $f(x)$.

$$n(x)=E[f(x)] \quad (12)$$

The specific expression is as follows.

$$v(x, x') = E[(f(x) - n(x)(f(x') - n(x'))] \quad (13)$$

Using Gaussian process to represent the unknown objective function, the marginal likelihood distribution can be obtained.

$$p(z \mid X, \delta) = \int p(z \mid f) p(z \mid X, \delta) df \quad (14)$$

where z is the set of observations $\{z_1, z_2, \ldots, z_m\}$. δ represents the hyperparameter, which is obtained by maximizing the marginal likelihood distribution. δ is the hyperparameter of the observed value.

2. Acquisition Function

The acquisition function is constructed with the posterior distribution of the observation dataset. The next evaluation point is selected by maximizing the acquisition function. The confidence boundary strategy is used as the acquisition function. The expression is shown in (15).

$$CBS(y, V) = \mu(y) + \alpha\sigma(y) \quad (15)$$

where V is the observed dataset at the current time instance. $\mu(y)$ and $\sigma(y)$ are the mean function and the covariance function of the joint posterior distribution. α is the adjusting parameter to select sampling points.

By maximizing the acquisition function, the next evaluation location is firstly selected. It evaluates the objective function value according to the selected point. Then the optimization expands the observed dataset. It modifies the prior distribution. Finally, the optimization obtains the optimal observation value through continuous iteration.

The goal of Bayesian optimization is to find the minimum value of the complex nonconvex function. To improve the discriminator ability of auxiliary classifier, the objective function is set as the opposite number. The generator hyperparameters are fixed in the optimization process. Only the discriminator hyperparameters are optimized. The performance of the generator and discriminator can be improved through the game training mechanism.

C. Self-attention Mechanism

GAN depends on the convolution layer to construct the network. After multiple convolution layers, the calculation efficiency of sample correlation is low. With the help of self-attention mechanism, a non-local module is introduced into the GAN convolution framework. The attention value among local regions is calculated, which increases the calculation efficiency of long and local correlation.

The self-attention module is in the middle of convolution layer. The input is the characteristic map of the upper convolution layer. The output is added to the original sample to obtain the input of the next layer. The feature x from the upper hidden layer is transformed into f and g, where $f(x)=W_f x$, $g(x)=W_g x$. Then the attention value is calculated.

$$\alpha_{j,i} = \frac{\exp(k_{ij})}{\sum_{i=1}^{N} \exp(k_{ij})} \tag{16}$$

$$k_{ij} = f(x_i)^{\mathrm{T}} g(x_j) \tag{17}$$

$$b_j = \sum_{i=1}^{N} \alpha_{j,i} q(x_i) \tag{18}$$

$$q(x_i) = W_q x_i \tag{19}$$

where $\alpha_{j,i}$ represents the attention to the ith region when generating the jth region. W_q is the learnable weight matrix, which is implemented by a 1×1 convolution operation. Finally, the output weight of the attention module is added to the original sample. Then the result is used as the input of the next hidden layer.

$$z_i = \lambda b_i + x_i \tag{20}$$

where z is the output of the self-attention layer. λ is initialized to 0.

The attention mechanism is shown in Figure 2. The features extracted from the previous convolution layer are passed through two 1 × 1 convolution kernels. Two feature spaces are obtained. Two feature matrixes are multiplied. The attention output matrix and the original convolution matrix are added. Finally, the feature map is produced. C, H and W represent the channel number, height and width of the characteristic graph respectively. ⊗ represents matrix multiplication. ⊕ represents matrix addition.

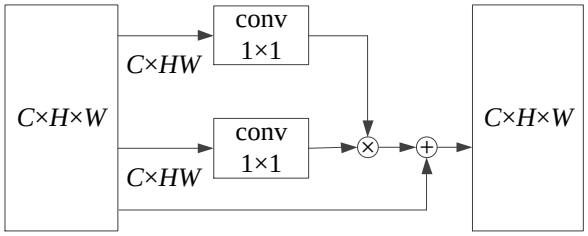


Fig.2 Attention mechanism

D. Discarding Method

Aiming at the over fitting problem for generation model training, the neuron discarding method is used. As a deep learning model, GAN contains many neurons and connection coefficients. If all connection coefficients are updated during the training process, the convergence speed of the parameter is slow. Therefore, the connection coefficients of some neurons are randomly discarded with a certain probability. A large number of neurons should be retained in the early training stage, which ensures the stability of model training and the diversity of the generated samples. The random discarding probability should be calculated as the following.

$$p(k) = b - \frac{a}{k+1} \tag{21}$$

where b is the probability of random discarding. Usually let b=0.5. a is the adjustment coefficient. Let a=0.3. k is the iteration number. The curve of equation (21) is shown in Figure 3. At the initial training stage, the discarding probability is small. With the increase of the iteration number, the discarding probability gradually tends to a constant value of 0.5. The original discarding method randomly discards half of the connected neurons in each layer. But the fixed percentage of discarding is not optimal. In Figure 3, the fixed ratio is changed to the decaying ratio. The decaying ratio has the advantage of preserving the original information, which ensures the stability of the model.

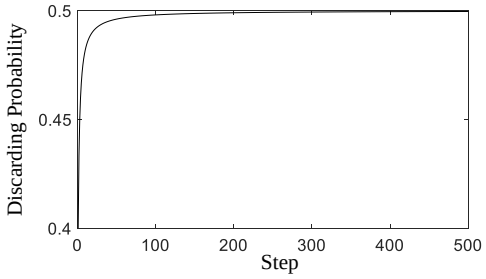


Fig.3 Random discarding probability

IV. Fault Diagnosis Process

The training steps of fault diagnosis based on ImGAN are as follows:

(1) Collect the original faulty signal. The dimension is divided reasonably.

(2) The ImGAN is modeled by the training set. The generator parameters are fixed. The discriminator parameters are trained.

(3) The discriminator parameters are iteratively optimized through the game mechanism. The model finally reaches the Nash equilibrium. The generator and discriminator are trained. The model parameters are saved.

(4) Under the condition of small samples, the generated samples are added to expand the training set. Then the fault diagnosis is carried out.

V. Experimental Design and Analysis

Tennessee Eastman Process (TEP) simulation data is used as the experimental dataset. In this experiment, three faulty classes are used. The dimension of each sample is 52. The dataset contains the measured variables and the manipulated variables. The number of training samples for each faulty class is 500. And the number of test samples for each faulty class is also 500.

A. Evaluating Indicator

To evaluate the model performance, the accuracy is selected as the evaluation index, as showed in formula (22). *TP*, *TN*, *FP*, *FN* are shown in Table 1.

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{22}$$

TABLE I CONFUSION MATRIX OF TRUE LABELS AND PREDICTIONS

True	Prediction	
	Positive	Negative
Positive	*TP* (True Positive)	*FN* (False Negative)
Negative	*FP* (False Positive)	*TN* (True Negative)

In this paper, the area under the ROC curve (AUC) is used to compare the accuracy difference between models. AUC measures the classification performance. The calculation formula of AUC is shown in the following formula.

$$AUC = \frac{\sum_{i=1}^{m_0} l_i - m_0 \times (m_0 + 1)/2}{m_0 \times m_1} \tag{23}$$

where m_0 and m_1 represent the number of negative and positive examples respectively. l_i is the ith ranking of negative examples among the whole testing samples.

B. Training Analysis

To compare the convergence performance of ImGAN with the prototype GAN, the experiment is carried out on the generator and discriminator. The generator score of GAN and ImGAN is shown in Figure 4. The score of the two networks decreases gradually during the whole training process. The fluctuation of the GAN score changes greatly. The GAN score is higher than the ImGAN score after 500 samples. It indicates that the ImGAN generator converges better. The experiment result reflects the continuous enhancement of the ImGAN generation ability. The discriminator score of GAN and ImGAN is shown in Figure 5. The score of the two networks increases gradually during the whole training process. The fluctuation of the GAN score changes greatly. The GAN score is lower than the ImGAN score after 500 samples. It demonstrates the continuous enhancement of the ImGAN discrimination ability.

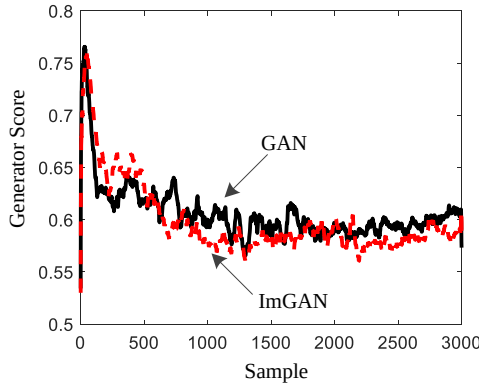


Fig.4 GAN and ImGAN generator score value trends

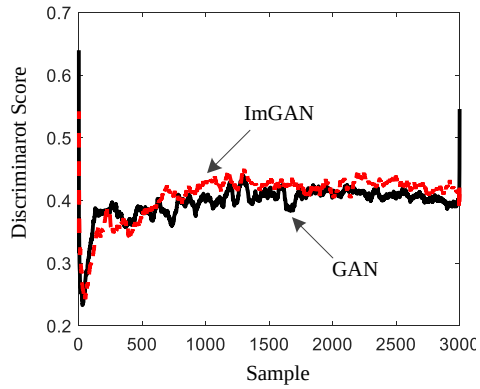


Fig.5 GAN and ImGAN discriminator score value trends

C. Authenticity Analysis of the Generated Sample

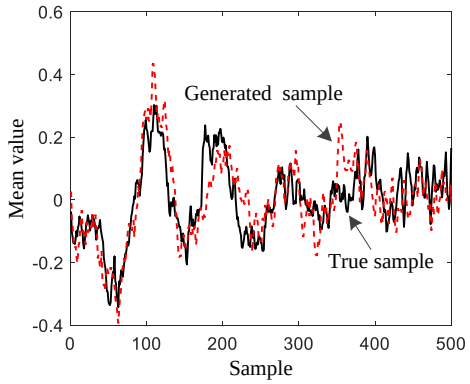
The value of average, variance and maximum is shown in equations (24)-(26). The statistical characteristic of the generated sample and the true sample is analyzed. $y(k)$ represents the sample sequence. The average value reflects the vibration range of faulty samples. The variance value reflects the dispersion of faulty samples. The maximum value reflects the vibration amplitude of the faulty sample. The probability density function (PDF) curve is used to show the probability distribution characteristic.

$$Mean = \frac{1}{M}\sum_{i=1}^{M} y(k) \tag{24}$$

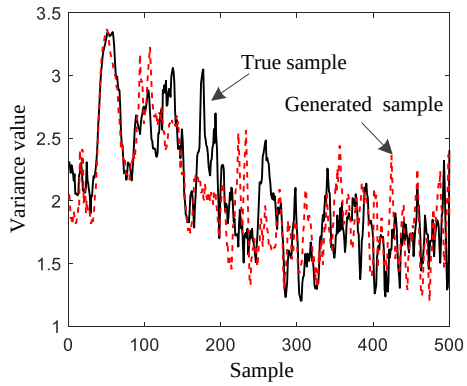
$$Var = \frac{1}{M}\sum_{i=1}^{M} \left(y(k) - Mean\right)^2 \tag{25}$$

$$Max = \max |\, y(k) \,| \tag{26}$$

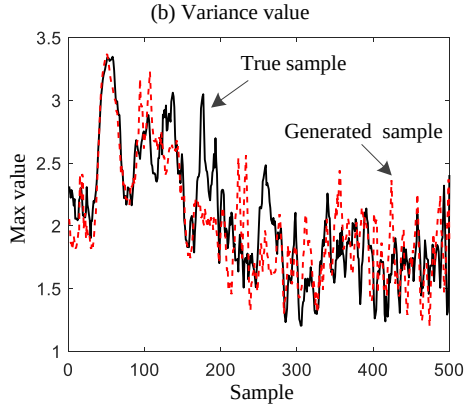
The curve diagram and PDF of the three statistics are shown in Figure 6 and Figure 7 respectively.



(a) Mean value



(b) Variance value

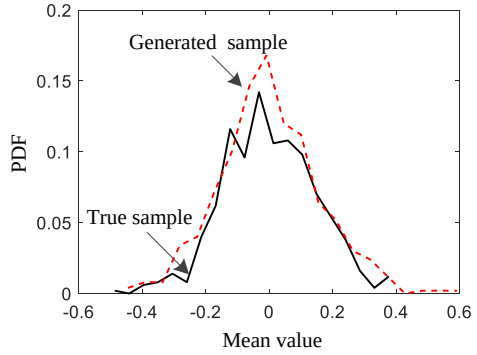


(c) Max value

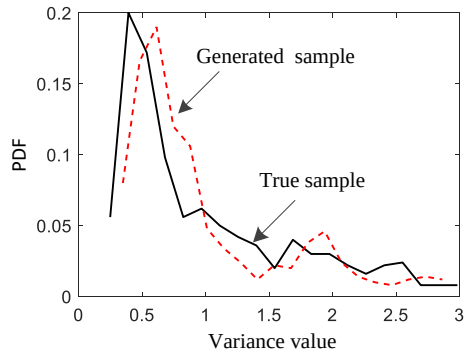
Fig.6 Statistics of true sample and generated sample

For Figure 6, the solid black line represents the distribution of real samples under different statistics. The dotted red line represents the distribution of the generated samples. For Figure 7, the solid black line represents the PDF curve of the

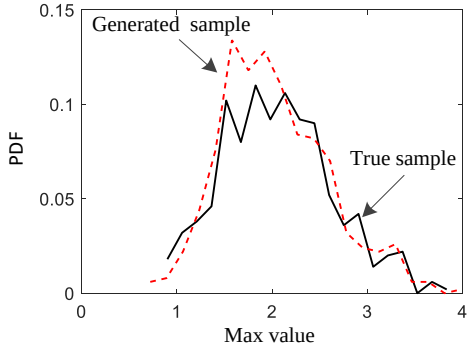
real sample under different statistics. The dotted red line represents the PDF curve of the generated sample under different statistics. Seen from the PDF curves of the three statistics, the probability distribution of the generated samples is close to that of the real samples. The generated sample line basically covers the real sample line, while the range is large. It is proved that the probability distribution of the generated samples is close to that of the true samples.



(a) PDF of mean value



(b) PDF of variance value



(c) PDF of max value

Fig.7 PDF of true sample and generated sample

D. Fault Diagnosis under Different Proportion of Small Samples

The random sample reduction method is adopted to reduce training samples by 30%, 50% and 70% respectively. The classification effect under different training samples is tested. The proposed method is analyzed in the small sample scenario. The construction of the training set is shown in Table 2.

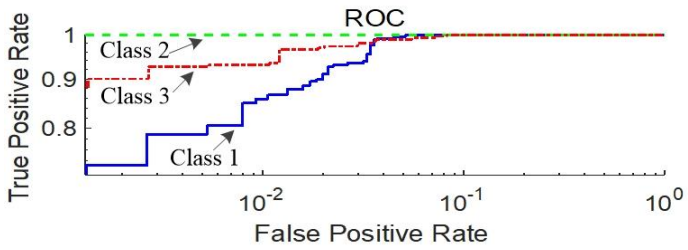
Taking the true positive rate as the y-axis and the false positive rate as the x-axis, the ROC curves under different training sets can be obtained. Area under curve (AUC) is defined as the area enclosed by the coordinate axis under the ROC curve. The area value is not greater than 1. The AUC value is equivalent to the probability of sorting the positive samples. The good performance will be obtained if the area of the ROC curve is close to 1. Under different training sets, the ROC curves are shown in Table 3 and Figure 8. The AUC of the proposed method is small under different training sets. When the training set is reduced by 70%, the AUC still reaches more than 0.99. The experiment result verifies the effectiveness of ImGAN in small sample scenario.

TABLE II TRAINING SAMPLE NUBER UNDER DIFFERENT REDUCTION

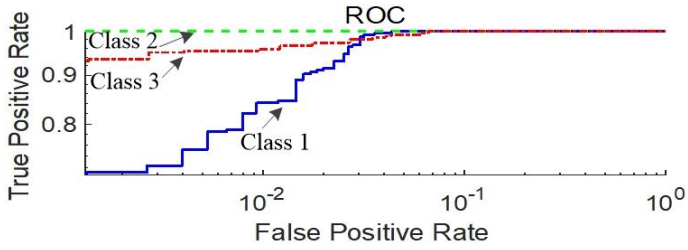
Training set size	Total number of training sample	Number of training sample for each class
30% reduction	1050	350
50% reduction	750	250
70% reduction	450	150

TABLE III AUC OF IMGAN CLASSIFICATION UNDER DIFFERENT TRAINING SET SIZE

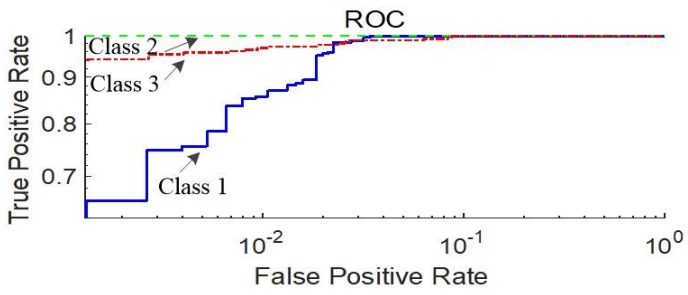
Training set size	Class 1	Class 2	Class 3
30% reduction	0.9987	1.0000	0.9962
50% reduction	0.9984	1.0000	0.9957
70% reduction	0.9964	0.9999	0.9945



(a) 30% reduction for training samples



(b) 50% reduction for training samples



(c) 70% reduction for training samples

Fig.8 Area under the curve of the ImGAN after reduction of training samples

E. Comparison of Different Classification Methods

To further verify the effectiveness and advancement of the proposed method, the comparison with SVM and CNN is

performed. The classification accuracy rate is taken as an index to verify the proposed method. Seen from Table 4, compared with SVM and CNN, the proposed method has the highest accuracy in different small sample scenario. When the sample reduction ratio is 70%, the average classification accuracy of SVM and CNN is 91.4% and 93.0%, respectively. While the classification accuracy of ImGAN is 97.6%. The comparison shows that the proposed method has the highest classification accuracy under different scenario, which proves the effectiveness of the proposed method.

TABLE IV FAULT DIAGNOSIS ACCURACY OF DIFFERENT METHODS

Method	30% reduction	50% reduction	70% reduction
SVM	97.8%	95.6%	91.4%
CNN	98.4%	96.6%	93.0%
ImGAN	98.6%	98.0%	97.6%

VI. Conclusion

To solve the problem of fault diagnosis under small sample, an improved fault diagnosis method based on GAN is established. Through the experimental verification, changing the structure of the generator and introducing Bayesian optimization can solve the problem of unbalanced dataset classification. The proposed method can be applied to fault diagnosis when it is difficult to obtain faulty samples.

References

[1] Cai G, Wu L, and Li M, "The Circuit Fault Diagnosis Method Based on Spectrum Analyses and ELM," *2021 IEEE 16th Conference on Industrial Electronics and Applications*, 2021.

[2] Yang P, Li Z, and Yu Y, "Studies on fault diagnosis of dissolved oxygen sensor based on GA-SVM," *Mathematical Biosciences and Engineering*, vol. 18, no. 1, pp386-399, 2021.

[3] Jin T, Yan C, and Chen C, "New domain adaptation method in shallow and deep layers of the CNN for bearing fault diagnosis under different working conditions," *The International Journal of Advanced Manufacturing Technology*, vol. 10, no. 10, pp1-12, 2021.

[4] Wen Yunfen, Zhao Rongzhen, and Xiao Youqiang, "Frequency safety assessment of power system based on multi-layer extreme learning machine," *Automation of Electric Power System*, vol. 43, no. 1, pp133-140, 2019. (in Chinese)

[5] Liu Yongqia, Shao Zhenzhou, and Wang Zheng, "Piecewise support vector machine model for theoretical wind-power calculation," *Acta Energiae Solaris Sinica*, vol. 31, no. 3, pp673-680, 2019. (in Chinese)

[6] Li Zhong, Zhang Weihua, and Sun Na, "Calculation of vibration fundamental frequency amplitude of transformer surface based on generalized regression neural network," *High Voltage Engineering*, vol. 43, no. 7, pp2287-2293, 2017. (in Chinese)

[7] Zhou P, Hu X, and Li P, "Online feature selection for high dimensional class-imbalanced data," *Knowledge Based Systems*, vol. 136, no. 1, pp187-199, 2017.

[8] Quan Z, Maozu G, and Yang L, "A classification method for class imbalanced data and its application on bioinformatics," *Journal of Computer Research and Development*, vol. 47, no. 8, pp1407-1414, 2010.

[9] Hao C A, Cl A, and Wy B, "Deep balanced cascade forest: An novel fault diagnosis method for data imbalance," *ISA Transactions*, 2021.

[10] Peng C, Li L, and Chen Q, "A Fault Diagnosis Method for Rolling Bearings Based on Parameter Transfer Learning under Imbalance Data Sets," *Energies*, vol. 14, 2021.

[11] Zhang J, Zhang Q, and He X, "Compound-Fault Diagnosis of Rotating Machinery: A Fused Imbalance Learning Method," *IEEE Transactions on Control Systems Technology*, vol. 99, pp1-13, 2020.

[12] Girija Attigeri, Manohara Pai M M, and Radhika M Pai, "Supervised Models for Loan Fraud Analysis using Big Data Approach," *Engineering Letters*, vol. 29, no. 4, pp1422-1435, 2021.

[13] S.F. Woon, S. Karim, M.S.A Mohamad, L. Ryan, and V. Rehbock, "On the Modification of the Discrete Filled Function Algorithm for Nonlinear Discrete Optimization," *IAENG International Journal of Applied Mathematics*, vol. 51, no. 4, pp930-935, 2021.

[14] Goodfellow I J, Pouget-Abadie J, and Mirza M, "Generative Adversarial Networks," *Advances in Neural Information Processing Systems*, vol. 3, pp2672-2680, 2014.

[15] Zheng C, Koh V, and Bian F, "Semi-supervised generative adversarial networks for closed-angle detection on anterior segment optical coherence tomography images: an empirical study with a small training dataset," *Annals of Translational Medicine*, vol. 9, no. 3, pp1-5, 2021.

[16] Tao C, Wang H, and Qi J, "Semi-Supervised Variational Generative Adversarial Networks for Hyperspectral Image Classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 90, pp1-1, 2020.

[17] Anokye F, and Kahanda I, "BioSGAN: Protein-Phenotype Co-mention Classification Using Semi-Supervised Generative Adversarial Networks," *2021 IEEE 34th International Symposium on Computer-Based Medical Systems*, 2021.

[18] Hirofumi Miyajima, Noritaka Shigei, Hiromi Miyajima, and Norio Shiratori, "Federated Learning with Divided Data for BP," *Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists* 2021, 20-22 October, 2021, Hong Kong, pp94-99.

Li Dongping is an associate professor in the Institute of Information Engineering of Kunming University. Her major is computer science and technology engineering. Her research direction includes big data technology and data mining.

Yang Yingchun is a senior engineer in Yunnan Branch of China Telecom Co., Ltd. His major is computer science and technology. His interest is big data analysis and algorithm research.

Shen Shikai is a professor in the Institute of Information Engineering of Kunming University. His research direction includes mobile communication network performance analysis, sensor network node location, data fusion theory and technology research.

He Jun is a professor in the Institute of Information Engineering of Kunming University. His major is computer software engineering. His research direction includes machine learning, data analysis and software evolution.

Shen Haoru is a lecturer in the Institute of Information Engineering of Kunming University. His major is big data research of computer networks and education.

Yue Qiang is an associate professor in the Institute of Information Engineering of Kunming University. His major is software engineering. His research direction includes big data technology and data mining.

Hong Sunyan is a lecturer in the Institute of Information Engineering of Kunming University. Her research direction includes machine learning algorithm.

Deng Fei is a senior engineer in the Institute of Information Engineering of Kunming University. His research direction includes artificial neural network and data mining.