

Research Article

Rotating Machinery Fault Diagnosis Method Based on Improved Semisupervised Generative Confrontation Network

Lei Wang¹,^{ID} Qian Li,¹ and Jin Qin²

¹Guizhou Health Vocational College, Tongren, Guizhou 554300, China

²Guizhou University, Guiyang, Guizhou 550000, China

Correspondence should be addressed to Lei Wang; dc-wang6688@st.btbu.edu.cn

Received 28 September 2021; Accepted 23 November 2021; Published 23 December 2021

Academic Editor: Shujaat Hussain Kausar

Copyright © 2021 Lei Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Error diagnosis and detection have become important in modern production due to the importance of spinning equipment. Artificial neural network pattern recognition methods are widely utilized in rotating equipment fault detection. These methods often need a large quantity of sample data to train the model; however, sample data (especially fault samples) are uncommon in engineering. Preliminary work focuses on dimensionality reduction for big data sets using semisupervised methods. The rotary machine's polar coordinate signal is used to build a GAN network structure. ANN and tiny samples are utilized to identify DCGAN model flaws. The time-conditional generative adversarial network is proposed for one-dimensional vibration signal defect identification under data imbalance. Finally, auxiliary samples are gathered under similar conditions, and CCNs learn about target sample characteristics. Convolutional neural networks handle the problem of defect identification with small samples in different ways. In high-dimensional data sets with nonlinearities, low fault type recognition rates and fewer marked fault samples may be addressed using kernel semisupervised local Fisher discriminant analysis. The SELF method is used to build the optimum projection transformation matrix from the data set. The KNN classifier then learns low-dimensional features and detects an error kind. Because DCGAN training is unstable and the results are incorrect, an improved deep convolutional generative adversarial network (IDCGAN) is proposed. The tests indicate that the IDCGAN generates more real samples and solves the problem of defect identification in small samples. Time-conditional generation adversarial network data improvement lowers fault diagnosis effort and deep learning model complexity. The TCGAN and CNN are combined to provide superior fault detection under data imbalance. Modeling and experiments demonstrate TCGAN's use and superiority.

1. Introduction

With the rapid development of industrial technology and science and technology, rotating machinery is widely used in modern industrial fields such as electric power, aerospace, metallurgy, wind power, nuclear power, and national defense. At the same time, with the high speed, continuity, and automation in the operation of mechanical equipment, once the core components of the equipment such as rotors and bearings break down or fail to work, it will affect the normal operation of the entire mechanical system and even lead to its paralysis, resulting in inaccessibility estimated loss [1]. In the past few years, various accidents at home and abroad have shown the harm of mechanical equipment failure

without exception (Figure 1). For example, in 1988, the rotor speed failure of the No. 5 steam turbine unit of Qin Ling Power Plant caused economic losses of up to 100 million yuan; in 2000, the French “Concorde” passenger plane caught fire on its wing during take-off, which eventually caused an explosion, causing the death of 109 people. In 2011, a power system failure in Chile caused the capital Santiago and most of the surrounding areas to become black, affecting nearly 10 million people. [2] Therefore, carrying out effective state monitoring and fault diagnosis analysis of mechanical equipment can timely grasp the operating status of the equipment and prevent the occurrence of key equipment failures, which is of great significance for ensuring the safe and reliable operation of mechanical



FIGURE 1: Repair of rotating machinery failure.

equipment and avoiding major accidents and huge property losses [3].

Traditional fault diagnosis starts from data. First, the vibration acceleration signal of rotating machinery and its components is collected in the laboratory or industrial site, and then the signals are analyzed in the time domain, frequency domain, and time-frequency domain. These methods include short-time Fourier transform, wavelet transform, and Hilbert yellow transform [4] and finally use specific classifiers such as support vector machine and artificial neural network for pattern recognition, so as to achieve the purpose of fault diagnosis for rotating machinery. In recent years, due to the continuous development of deep learning and its excellent feature extraction capabilities, the use of deep learning for fault diagnosis has become one of the most popular breakthrough technologies. The deep learning model represented by convolutional neural networks has been widely used in the field of rotating machinery fault diagnosis [5]. However, regardless of whether it is a classic fault diagnosis method or a deep learning method, the field of rotating machinery fault diagnosis still faces some challenging problems. In actual production, the occurrence of rotating machinery faults is often random, and the collection of fault signals is more difficult and the amount of data is small. This resulted in insufficient samples and unbalanced data during fault diagnosis. To perform fault diagnosis of rotating machinery in the case of small samples and unbalanced data has gradually attracted the attention and discussion of the scientific research community [6].

2. Related Work

Due to the unstable dynamic system, noise, and modulation of rotating machinery equipment, its signal generally has the characteristics of nonlinearity and nonstationarity [7]. The signals collected directly during the operation of rotating machinery are generally one-dimensional vibration signals. The corresponding relationship between the signal and the

fault cannot be directly seen through the vibration signal, so a series of time-domain and frequency-domain analysis methods have been produced. Simply analyzing the signal in the time domain or the frequency domain cannot completely and accurately express the fault characteristics of the signal, and the analysis method based on the time-frequency domain can describe the time-varying characteristics and energy distribution of the signal, which is a more complete and accurate expression method. Therefore, the feature extraction method based on the time-frequency domain has become a hot issue in the field of fault diagnosis [8].

In the time-domain analysis, statistical methods were initially used to calculate the specific parameters of the signal, including the kurtosis of the signal, the root mean square value, the waveform index, and the peak index [9]. In addition, time-domain analysis also includes time-domain correlation analysis of signals, such as autocorrelation analysis and cross-correlation analysis. Only time-domain analysis of signals cannot meet engineering and production requirements, and these indicators are often not stable enough under complex working conditions and severe external interference. Therefore, in this case, the application of frequency-domain analysis is born, and the frequency-domain analysis of the signal can reflect the frequency-domain components and distribution of the signal, which has a good engineering and scientific research application prospect. The frequency-domain analysis of the signal mainly includes two kinds of spectrum analysis: classical spectrum and modern spectrum [10].

Blackman Turkey proposed the classic spectrum in 1958 [11], which is derived from the Fourier transform. Classical spectrum analysis methods include envelopment analysis (EA), zoom spectrum analysis (ZPA), etc. Classical spectrum analysis methods are very important for fault diagnosis of rotating machinery. EA is currently the most widely used diagnostic method in the field of spectrum analysis. The core idea of EA is to perform low-pass filtering on fault vibration signals. Mcfadden first applied EA technology in the field of rolling bearing fault diagnosis [12], and Li et al. completed

the noise reduction of gear vibration signals through EA [13]. The core idea of ZPA is to improve the resolution of some sensitive frequencies in the frequency domain through the application of filtering technology and resampling technology. Wang et al. found the characteristic frequency of the corresponding fault of the rolling bearing signal through ZPA and realized a good diagnosis of the rolling bearing fault [14]. Although classical spectrum analysis has been widely used, it still has some shortcomings, such as the requirement that the signal must be stable and cannot be applied to nonstationary signals.

When the classical spectrum analyzes the signal selected by the window function, it is assumed that the signal outside the window function is zero. This assumption is the main reason for the poor quality of the classical spectrum estimation. Modern spectroscopy implements a certain expectation or outlook outside the signal selected by the window function, which improves the quality of spectrum estimation [15]. The parametric model method is the core method of modern spectroscopy. Since E. Parzen proposed an autoregressive model spectrum estimation method in 1968, a series of spectrum estimation methods such as harmonic analysis method, maximum likelihood method, and auto-moving regression average method have gradually emerged [16].

After years of development, the pattern recognition of rotating machinery has made considerable progress. The most widely used applications in the field of fault recognition include support vector machines, semisupervised dimensionality reduction, and deep learning.

SVM is mainly used to deal with statistical classification and regression analysis problems, which is a kind of supervised learning algorithm. SVM adopts the mechanism of kernel function. A suitable kernel function is particularly important for SVM, which will affect the classification performance of SVM. The application of SVM in the fault diagnosis of rotating machinery has been very common, but SVM lacks the ability of feature extraction, so in the process of fault diagnosis it is often used in combination with signal processing methods. Cheng Junsheng et al. used EMD envelope spectrum to extract signal features and used SVM for classification to realize the fault diagnosis of rolling bearings [17]. Xu and Si applied an improved particle swarm algorithm to adjust the least squares SVM and achieved a good diagnostic effect for rolling bearings [18]. Li and Shu proposed the use of fuzzy clustering and complete binary tree combined with support vector machines to perform transformer fault diagnosis [19].

Many scholars introduced semisupervised dimensionality reduction into fault diagnosis and achieved good dimensionality reduction, fault identification, and classification effects. Li Chengliang used the marked sample information in the training set to establish constraint point pairs and combined the structural information of the sample connection graph to design the kernel function and proposed rotor system fault diagnosis based on semisupervised spectral kernel clustering to solve the problem. Jiang Li proposed a fault diagnosis model based on semisupervised Laplacian feature mapping in order to overcome the

problem of insufficient labeled samples in the fault data set with nonlinear characteristics and high dimensionality. Jiang et al. proposed a feature extraction method based on semisupervised kernel edge Fisher analysis and applied it to bearing fault diagnosis in order to extract the best features from the nonstationary and nonlinear fault mechanical vibration signals and improve the classification accuracy. Gao Zhiyong et al. proposed a chemical system state monitoring method based on improved local linear discriminant analysis. This method uses the label information of training samples to reconstruct the local interclass divergence matrix and improve the performance of the algorithm. Yi Weilin et al. proposed a reconstructed semisupervised ELM method based on the traditional ELM method and applied it to fault diagnosis. Fang Liqing et al. proposed a semisupervised neighborhood adaptive LLTSA algorithm for fault diagnosis in order to extract a subset of sensitive features with high recognizability. Luo et al. proposed a fault diagnosis method based on semisupervised manifold learning and conversion support vector machine for the scarcity of labeled samples in training samples. This method can achieve high fault diagnosis accuracy when the labeled samples are insufficient. Zhao Xiaoli et al. proposed a rolling bearing fault diagnosis method based on the maximum boundary projection dimension reduction of the regularization kernel to solve the problem of difficulty in obtaining fault samples in fault diagnosis.

Deep learning has developed rapidly in these years. The classic deep learning model is a perceptron with multiple hidden layers. The development results of deep learning have been widely used in the field of fault diagnosis of rotating machinery. According to the application of structure and technology, deep learning is mainly divided into three categories: generative deep structure, discriminative deep structure, and hybrid structure [20]. A deep belief network (DBN) is one of the typical generative deep structures, and its composition structure is a series of restricted Boltzmann machine (RBM) units. The DBN solves the problem of slow convergence speed when training multiple layers of the traditional BP network. Li Weihua and others used the DBN to carry out bearing fault and identification and achieved good results [21]. Zhang Na et al. used the global dynamic learning rate to improve the DBN and applied it to the life prediction of rolling bearings [22]. A convolutional neural network is one of the typical discriminative deep structure models. The CNN can adaptively extract signal features and classify samples. Duy-TangHoang et al. directly used the bearing vibration signal as the input of the neural network and used the CNN to realize the fault identification of the rolling bearing and achieved a good diagnosis effect [23]. Luo et al. proposed a method of combining discrete wavelet time-frequency transform and convolutional neural network for rolling bearing fault diagnosis [24]. Jiang and Wang proposed AFDCNN based on the adaptive Fisher criterion and applied it to the quantitative diagnosis of gear faults [25]. The hybrid structure model usually aims to be discriminative but uses the output of the generative structure to make the goal easier to optimize. Jiang Yunjie and others used BP networks to optimize DBNs to achieve end-to-end

situation assessment. A generative adversarial network (GAN) is a new deep learning model that has emerged in the past two years. It can be used for sample generation or feature segmentation. It has been widely used in the field of speech and image. Zhu Chun et al. used deep convolutional generative adversarial networks (DCGANs) to realize the autonomous generation of speech signals. Jiang Yun et al. used conditional generation adversarial networks to segment the bite wing image and achieved good results. In addition, generative countermeasure networks are gradually being applied to the field of fault diagnosis [26]. Dai Jun et al. combined generative countermeasure networks and autoencoders to detect abnormalities in mechanical systems, achieving faster and more accurate detection of system abnormalities.

3. Method

3.1. Overview of Semisupervised Dimensionality Reduction Methods. In the actual application of supervised dimensionality reduction algorithms, labeled samples are usually difficult to obtain and have a small number, while it is relatively easy to obtain unlabeled samples. Compared with traditional dimensionality reduction methods, semisupervised dimensionality reduction methods (SDA, SELF, etc.) can use the data category labeling information and the structure information of all sample data at the same time, so that the data between the classes after dimensionality reduction are scattered and within the class, while keeping its own covariance or local structure information unchanged, and a better classification and recognition accuracy rate can be obtained. Under normal circumstances, the establishment of the semisupervised dimensionality reduction method depends on the following three basic model assumptions, namely the semisupervised smoothness assumption, clustering assumption, and manifold assumption:

Semisupervised Smoothness Assumption. Assume that two neighboring samples in a dense data area have similar class labels, that is, when two samples are connected by edges in the dense data area, they have a high probability of having the same class label; on the contrary, when two samples are distinguished by sparse data, they tend to have different classification standards.

Cluster Assumption. When two samples are in the same cluster, they have a greater probability of having the same class label. Its equivalent is defined as the hypothesis of low-density separation, that is, the classification decision boundary should pass through the sparse sample area as much as possible, so that data points with a greater probability of the same category are classified on the same side of the decision boundary, avoiding being located in a dense area the samples are divided into two sides of the decision boundary.

Manifold Assumption. When embedding high-dimensional data into a low-dimensional manifold, two samples located in the same local neighborhood in the low-dimensional manifold have more similar class labels.

3.1.1. Semisupervised Discriminant Analysis Algorithm. The LDA algorithm is a supervised dimensionality reduction algorithm that extracts features from the space by maintaining the interclass dispersion and the intraclass dispersion. However, in practice, due to the small number of labeled training samples, the covariance matrix of each category may not be accurately estimated. Therefore, the generalization ability of the test sample cannot be guaranteed. One possible way to solve the shortage of training samples is to learn both labeled and unlabeled data. Cai et al. proposed semisupervised discriminant analysis by adding regularization items to LDA. The SDA algorithm can use label samples to maximize the separability between different categories. Unlabeled samples can be used to maximize the preservation of the internal geometric structure of the data and overcome the small sample problem in LDA.

Suppose the original data set $X = (x_1, x_2, \dots, x_n) \in R$ where the labeled sample set is $X_l = (x_1, x_2, \dots, x_l)$ and the unlabeled sample set is $X_u = (x_{l+1}, x_{l+2}, \dots, x_{l+u})$, that is, $X = X_l \cup X_u$. The projected result of the data set is $Y = (y_1, y_2, \dots, y_n)$.

The objective function of LDA is defined as

$$\arg \max_T \frac{T^T S^{(b)} T}{T^T S^{(t)} T}. \quad (1)$$

Without loss of generality, assuming $m = 0$, then

$$S^{(b)} = \sum_{l=1}^c n_l m_l m_l^T = \sum_{l=1}^c n_l \left(\frac{1}{n_l} \sum_{i=1}^{n_l} x_i^{(l)} \right) \left(\frac{1}{n_l} \sum_{i=1}^{n_l} x_i^{(l)} \right)^T = \sum_{l=1}^c X^{(l)} W^{(l)} (X^{(l)})^T, \quad (2)$$

where $X^{(l)} = (x_1^{(l)}, x_2^{(l)}, \dots, x_{n_l}^{(l)})$ means the data matrix of the i -th type, $W^{(l)}$ is the $n_l \times n_l$ matrix, all element values in $W^{(l)}$ are $1/n_l$, and n_l represents the number of samples of type l . Then, the interclass dispersion matrix can be transformed into

$$S^{(b)} = \sum_{l=1}^c X^{(l)} W^{(l)} (X^{(l)})^T = X W_{n \times n} X^T. \quad (3)$$

Therefore, the objective function of LDA can be equivalently transformed into

$$\arg \max_T \frac{T^T S^{(b)} T}{T^T S^{(t)} T} = \arg \max_T \frac{T^T X W_{n \times n} X^T T}{T^T X X^T T}. \quad (4)$$

To prevent the LDA algorithm from overfitting when the labeled samples are insufficient, the regularization term $J(T) = 2TTXLXTT$ is introduced into the objective function of LDA (by constructing the k-nearest neighbor graph to maintain the manifold of the data), so that SDA The maximum dispersion between classes and the local structure of the data can be maintained at the same time. The objective function of the SDA method is

$$\max_T \frac{T^T S^{(b)} T}{T^T S^{(t)} T + \alpha J(T)} = \max_T \frac{T^T S^{(b)} T}{T^T (S^{(t)} + \alpha XLX^T) T}, \quad (5)$$

where $a \in [0, 1]$ is the regularization parameter, and L is the Laplacian matrix. Solve the following generalized eigenvalue problem:

$$S^{(b)} a = \lambda (S^{(t)} + \alpha XLX^T) a. \quad (6)$$

The eigenvectors corresponding to the first d largest eigenvalues are selected to form the best conversion matrix. The original high-dimensional data X is obtained by $Y = T^T X$ and the low-dimensional data Y after dimensionality reduction.

3.1.2. Semisupervised Local Fisher Discriminant Analysis (SELF) Algorithm. When LFDA has insufficient label samples, due to excessive reliance on a small number of label samples, it is easy to fall into learning. In order to solve this problem, Sugiyama et al. effectively merged LFDA and PCA and proposed a new semisupervised partial Fisher discriminant analysis (semisupervised local Fisher discriminant analysis) algorithm. This method combines the advantages of PCA and LFDA, comprehensively considers labeled samples and unlabeled samples, maximizes the distance between the data between classes after dimensionality reduction, minimizes the distance between adjacent data within the class, and keeps the non-intraclass as much as possible. The structure of the adjacent data overcomes the shortcomings of the unsupervised properties of the PCA method and the high-dimensional small sample problem in the LFDA method.

Suppose a data set $X = \{x_i \in R^D, i = (1, 2, \dots, n)\}$, X_l is label data, and X_u is unlabeled data, that is, $X = X_l$, combined with PCA and LFDA methods, SELF is obtained. The objective function of is expressed as

$$T^{(SELF)} = \arg \max_{T \in R^{D \times d}} \frac{\text{tr}\{T^T S^{(slb)} T\}}{\text{tr}\{T^T S^{(slw)} T\}}, \quad (7)$$

where $S^{(slb)}$ is the regularized local interclass divergence matrix of SELF, and $S^{(slw)}$ is the regularized local intraclass divergence matrix of SELF, which are defined as follows:

$$\begin{aligned} S^{(slb)} &= (1 - \beta) S^{(lb)} + \beta S^{(t)}, \\ S^{(slw)} &= (1 - \beta) S^{(lw)} + \beta I_d, \end{aligned} \quad (8)$$

where $S^{(t)}$ is the global divergence matrix of PCA, defined as equation (9). I_d is the standard matrix, and $\beta \in [1, 0]$ is weigh parameters, which can make SELF have the characteristics of LFDA and PCA. It can increase the flexibility of the algorithm through adjusting β . Obviously, when $\beta = 0$, SELF degenerates into LFDA, and when $\beta = 1$, SELF is equivalent to PCA. The global divergence matrix of PCA is defined as

$$S^{(t)} = \frac{1}{2} \sum_{i,j} W_{i,j}^{(t)} (x_i - x_j)(x_i - x_j)^T. \quad (9)$$

The conversion matrix T in equation (7) can be solved by the generalized eigenvector calculation method shown in equation (10).

$$S^{(slb)} a = \lambda S^{(slw)} a, \quad (10)$$

where λ is the eigenvalue, select the eigenvectors corresponding $(a_1, a_2, \dots, a_{1d})$ of the first d largest generalized eigenvalues to form the transformation matrix T , and obtain the projection Y of the original high-dimensional data X in the low-dimensional space.

$$Y = T^T X. \quad (11)$$

The SELF method can use labeled samples to calculate the local intraclass divergence matrix and the local interclass divergence matrix, so they are not easily affected by the data distribution structure; at the same time, the SELF method can use the local class defined by the PCA method when the selected label is not ideal. The inner divergence matrix and the local interclass divergence matrix are used to find a better projection direction, so it has a better dimensionality reduction effect.

3.2. Rotating Machinery Fault Diagnosis Based on IDCGAN

3.2.1. Improved Deep Convolution Generative Adversarial Network. The activation function determines whether the signal can be transmitted in the DCGAN. The key to the neural network's ability to deal with nonlinear problems is that the activation function can store and express the "features of activated neurons" through the nonlinear function. In the DCGAN, the operation process of convolution and deconvolution is linear. If the activation function of nonlinear mapping is not used, then the DCGAN only has the function of linear expression. The use of the activation function can improve the expressive ability of the network, making the entire network into a nonlinear model, so that the DCGAN can solve complex nonlinear problems.

Common activation functions include sigmoid function, Tanh function, ReLU function, and LeakyReLU function. Sigmoid and Tanh are commonly used nonlinear activation functions. The sigmoid function can transform the continuous value of the input into an output between 0 and 1. It is the most common activation function used initially. However, it has some significant shortcomings, such as excessive calculation, easy gradient disappearance, and nonzero mean output. Tanh is an improvement of sigmoid,

and its output is zero mean, but the shortcomings of vanishing gradient and large amount of calculation have not been improved. The ReLU function effectively solves the problem of the disappearance of the gradient of the sigmoid function, and its calculation speed and convergence speed are much faster than the sigmoid function and the Tanh function. But using the ReLU function, when the forward propagation input is less than 0, the neuron will be in an inactive state and will “kill” the gradient during the backward propagation. The LeakyReLU function avoids the above-mentioned problems of the ReLU function by setting the negative semiaxis coefficient a . However, during the experiment, it was found that as the number of iterations increases, LeakyReLU may cause the DCGAN to oscillate, nonconvergence, or even overfitting, resulting in distortion of the generated samples.

In response to the above problems, this paper proposes a new activation function, which is defined as the IReLU function. On the basis of the LeakyReLU function, the IReLU function sets a condition for the input x . Instead of learning all the values, it passes a given threshold. When the input is greater than a specific value, it stops learning features. This reduces the possibility of network oscillations in the later training stage, alleviates the overfitting situation, and retains the advantages of the LeakyReLU function. When the value of x is taken, the function is no longer active and needs to be determined according to different data sets. Through many experiments on different data sets, it is found that the IReLU function not only improves the stability of DCGAN but also can produce higher-quality samples. Applying it to the fault diagnosis of rolling bearings can effectively improve the recognition rate of fault diagnosis. The mathematical expression of the IReLU function is as follows:

$$f(x) = \begin{cases} b, & x \geq b, \\ x, & 0 \leq x \leq b, \\ ax, & x < 0. \end{cases} \quad (12)$$

In the classic DCGAN, the activation function of the output layer of the generator is selected as Tanh, the activation functions of other layers are selected as ReLU, and the activation functions of all layers of the discriminator are selected as LeakyReLU. The specific process of implementing IDCGAN is as follows:

- (1) The preprocessing link converts the one-dimensional time series signal into a two-dimensional polar coordinate map and performs gray-scale and normalization processing.
- (2) Set the minimum batch size, and specify the network structure parameters of the generator and the discriminator in advance.
- (3) Keep the parameters of the discriminator unchanged and update the generator. Connect the generator and the discriminator; the generator receives the random noise x and generates a generated sample $G(x)$. Set the label of the generated sample to 1, input the

generated sample into the discriminator for forward and back propagation to obtain the partial derivative of the discriminator network parameters, use the partial derivative of the discriminator to find the partial derivative of the generator, and use the optimizer method. Only use the partial derivative of the generator to update the generator.

- (4) Keep the generator parameters unchanged and update the discriminator. The generator receives random noise x and generates a sample $G(x)$. Set the label of the generated sample to 0 and the label of the real sample to 1. Input the generated sample and the real sample into the discriminator for forward and back propagation to obtain the partial derivative of the discriminator's network parameters, and update the discriminator with the partial derivative of the discriminator in the manner of an optimizer.
- (5) Perform 1:1 iterative update, repeat steps 3 and 4, and continuously optimize the generator and discriminator parameters. After reaching the given number of iterations, the iteration stops.

4. Experiments and Results

4.1. Time Series Conditional Generation Confrontation Network Data Method. In the field of fault diagnosis of rotating machinery, the data collected in a factory or laboratory environment is generally vibration acceleration data with timing information. Usually only a small amount of labeled data can be used for training, and some fault conditions data are often especially lacking. By generating a confrontation network, data enhancement can be achieved and the problem of lack of samples can be solved. However, when using a generative countermeasure network, it is necessary to convert a one-dimensional time series signal into a two-dimensional image signal. This increases the workload of fault diagnosis, and also the generative confrontation network suitable for two-dimensional images often has a complicated model and takes too long to train. This chapter directly starts from the time series vibration acceleration data of the original rotating machinery and proposes to use the time series condition to generate the confrontation network to directly enhance the vibration data. The experimental results prove the feasibility and superiority of TCGAN.

Many data enhancement techniques have been applied to the image field, including image inversion, translation, and rotation. However, none of these methods can be used well in time series data enhancement. For time series data, it is impossible to confirm whether these transformations have changed the nature of the time series through simple visual comparison. The current mainstream time series data augmentation technology mainly includes two methods: data slicing and data deformation.

Data slicing technology is a technology inspired by computer vision. The method mainly includes two parts: cropping time series slices and classifying the cropped slices. In the training process, the classifier needs to learn the

classification information of each slice, and the size of the slice is an important parameter of the data slice. In the test process, the predicted label is determined by classifying the fragments obtained in the time series and voting on all the labels. When selecting time slicing for data enhancement, the effect may be unsatisfactory because cutting the time series often eliminates the time correlation in the data, resulting in a decrease in classification accuracy.

Data deformation is another data enhancement method, and slice size and warpage rate are important parameters. It achieves the purpose of expanding the data set by stretching randomly selected fragments in the time series. In theory, the premise of using data deformation is that this method does not change the distribution of training data. But in some practical situations, the time scale is of great significance. When data deformation is used, the data may have different interpretations, and that is why the data deformation cannot be well promoted.

4.2. Principle and Structure of Time Series Generation Adversarial Network. In order to solve the problem of fault diagnosis under the imbalance of one-dimensional signal data of rotating machinery, this chapter proposes a time series conditional generation countermeasure network. Given a periodic sampling time data set S , the goal of data enhancement is to generate a new time series that has the same properties and distribution as the time series in the data set S . In order to achieve this goal, this chapter designs a time-aware conditional generative confrontation network,

which adjusts the generator and discriminator through time steps. The goal of TCGAN is to obtain the latent spatial distribution of the time series to achieve the goal of imitating the dynamics of the time series. Because this chapter is for the rotating machinery data set, it is assumed that the time series are noisy.

GAN was proposed by I. Goodfellow as a generative model. The GAN model includes two parts: generator G and discriminator D . The generator is used to learn the distribution of real samples, and the goal of the discriminator is to determine whether the data come from the training data or the generator. The generator establishes a mapping function between the prior noise distribution $p_z(z)$ and the training data set, so as to learn the same sample distribution p_g as the training data set. The discriminator essentially consists of two classifiers, and the output of the discriminator is scalar, indicating whether x is from the training set or the p_g probability. The generative confrontation network can be extended to a conditional generative confrontation network. If both the generator and the discriminator are conditioned on other additional information y , y is used as an additional feedback input to the generator and the discriminator. In the generator, the prior noise $p_z(z)$ and y are combined into a hidden representation, and the adversarial training framework makes this hidden representation very flexible. In the discriminator, c and y are used as the input of the discriminator. At this time, the objective function of the conditional generation confrontation network can be expressed in the following form:

$$\min_G \max_D V(D, G) = E_{X \sim P_{\text{data}}(X)} [\log D(x)] + E_{Z \sim P_Z(Z)} [\log (1 - D(G(z)))]. \quad (13)$$

On the basis of the above principles, this chapter proposes the TCGAN model. The TCGAN is composed of two convolutional neural networks, which constitute a generator and a discriminator. Z is the noise space used to generate the model samples, and $z \in Z$ is the sample collected from the prior noise distribution $p_z(z)$. T is used to adjust the time

step space of the generator and the discriminator, and X represents the time series data space of the generator output or the discriminator input. When training data or conditional data related to it is used, the density model is defined as $P_{\text{data}}(x, t)$.

$$\min_G \max_D V(D, G) = E_{X \sim P_{\text{data}}(X)} [\log D(xt)] + E_{Z \sim P_Z(Z)} [\log (1 - D(G(zt)))]. \quad (14)$$

Among them, $t = t_1, \dots, t_n$ is the time step vector obtained by sampling and sorting from T . The TCGAN model can generate new time series data corresponding to no time step in the training data.

The generator performs the function $G(Z, T) \rightarrow X$, where $t \in T$ and $z \in Z$ are the input data, $x \in X$ is the generated time series, the generator X is a convolutional neural network, which takes the input noise and time step as the input of the model, and the output is the corresponding time step. This transformation is accomplished through four deconvolution layers with ReLU activation functions, and

batch normalization is performed on each layer except the last layer. The goal of the generator is to adjust its corresponding parameter $\log(1 - D(G(zt)))$ to minimize, $z \in Z$ is the noise vector, and $t \in T$ is the time step vector.

The discriminator performs the function $D(X, T) \rightarrow [0, 1]$, D takes the actual real data or generated data and their corresponding time steps as input, and the output is a value between 0 and 1, where 0 represents the generated time series, and 1 represents the real data. The discriminator is composed of two convolutional layers. After each convolutional layer, there is a maximum pooling layer,

and there is a fully connected layer at the end of the network. The goal of the discriminator is to adjust its corresponding parameter $\log D(xt)$ to maximize, where x is the time series vector and t is the time step vector. Figure 2 shows the network structure of TCGAN's generator and discriminator.

4.3. Experimental Comparison and Analysis. In order to verify the impact of TCGAN on the classification accuracy of unbalanced data sets, this chapter first constructs a simulation composite data set. The data consist of two types of signals: sine wave and sawtooth wave. The parameters used to participate in the training include the number of curves S for training TCGAN and the length of the signal time series L .

In order to evaluate the pros and cons of the TCGAN simulation signal sample expansion quality, we choose to use convolutional neural network to classify the signal. As shown in Table 1, the CNN model trained on the TCGAN extended data and the real data are trained on the CNN model, and the real data are used for testing to achieve model performance comparison. The CNN recognition rate of the two models is counted and the AUROC curve is drawn.

It can be seen from the table that when the number of training signals N is between 40 and 100, and the signal length L is between 40 and 100, the real test signal is compared with the real CNN model trained by the TCGAN extended signal. The classification accuracy of the signal training model is not much different, all reaching more than 95%. This directly shows that the model trained by the generated signal and the model trained by the real signal have the same recognition performance in CNNs and indirectly shows that the effect of TCGAN's simulation signal sample generation is good.

The AUROC curves of the two models with different numbers of training signals are shown in Figures 3–6. It can be seen from the figures that when the number of training signals is 40, 60, 80, and 100, the classifier curve trained by the TCGAN generator and the classification of the real signal training the AUROC curve are basically the same. The AUROC values are all above 0.95, which directly indicates that the performance of Model 1 and Model 2 classifiers is equivalent and indirectly indicates that the TCGAN has a good effect in generating simulation signals and can replace real signals to train CNN classifiers.

In order to further verify the effect of TCGAN on the performance improvement of the classifier under the imbalanced data set, when N is 100 and the signal length is 40, the imbalance rate of the data is gradually increased. The AUROC results are shown in Figures 7 and 8. It can be seen from the table that when the imbalance rate of the data gradually changes from 0.2 to 1 after the expansion of TCGAN, the AUROC value gradually increases, which further shows that TCGAN is important for improving the performance of the classifier.

Collect the vibration signal of the bearing through LMS. Using EDM technology, the individual faults of the inner ring, outer ring, and rolling elements are arranged on SKF6206-2RS1/C3 deep groove ball bearings. The loading load is 2 kN, and the spindle speed is 1500 r/min. The

sampling frequency is 4096 Hz, and every 200 points is a piece of data. 200 sets of data are collected for normal bearings, and 130 sets of data are collected for each of the three types of failures. Among them, 100 groups of normal bearings and 30 groups of three types of fault conditions are the classifier training set, and the remaining 100 groups of signals of each type are used as the classifier test set.

In order to evaluate the pros and cons of TCGAN's generated time sequence signal, time-domain feature extraction was performed on the generated signal and the real signal, and the differences between the two were compared.

In order to quantify the generation effect of TCGAN, we choose to compare the time domain and frequency domain characteristics of the two. Take a faulty outer ring bearing as an example, and the results are shown in Table 2. It can be seen from the table that the signal generated by the TCGAN is not much different from the real signal regardless of the time domain or frequency domain characteristics. Therefore, it can be shown that the TCGAN-generated signal has a good effect and has learned the characteristics of the real signal.

In order to verify the effect of TCGAN on the fault diagnosis of rotating machinery under the unbalanced data set, the model shown in Table 3 is designed for comparison. Model 1 is composed of real samples and simulates the fault diagnosis of imbalanced data sets in the case of small samples. Model 2 is composed of real samples and TCGAN-generated samples to generate samples instead of real samples to construct a balanced data set and compare the CNN recognition of the two models' rates.

The experimental results are shown in Table 4. It can be seen from the table that when there is no data enhancement, the CNN recognition rate is 65.5% when performing fault diagnosis on the unbalanced data set, and when the TCGAN sample is expanded, the imbalance of the data disappears. The CNN recognition rate is increased to 87.2%. This shows that TCGAN can well solve the problem of rolling bearing fault diagnosis in the case of data imbalance.

In order to further verify the improvement effect of TCGAN on the fault diagnosis of rotating machinery under the unbalanced data set, the data of gear cracks and broken teeth in the laboratory are selected for experiments. The gear speed is 360 r/min, and every 200 points are used as a segment of the signal, 30 sets of gear crack data and 40 sets of broken tooth data are collected, and the signal sampling frequency is 1024 Hz. Choose TCGAN data expansion for gear crack data. 10 sets of gear crack data are used as the input data of TCGAN.

From a qualitative point of view, TCGAN-generated signals basically learned the shape features of real samples. In order to quantify the generation effect of TCGAN, we choose to compare the time domain and frequency domain characteristics of the two. Taking the cracked gear as an example, the results are shown in Table 5. It can be seen from the table that the signal generated by TCGAN is not much different from the real signal regardless of the time domain or frequency domain characteristics. Therefore, it can be shown that the TCGAN has a good signal generation effect and has learned the characteristics of the real signal.

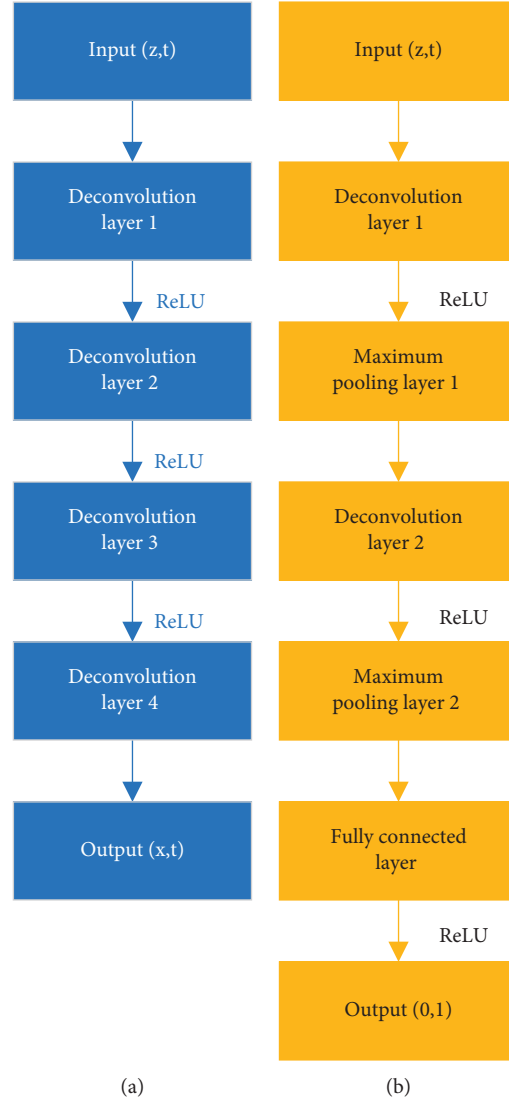


FIGURE 2: TGGAN structure schematic diagram: (a) builder and (b) discriminator.

TABLE 1: Simulation signal CNN experiment comparison.

Model	CNN training sample	CNN test sample
Model 1	Real sin and saw signal ($2 * S$)	Real sin and saw signal (S)
Model 2	Real sin and saw signal ($2 * S$)	Real sin and saw signal (S)

In order to verify the effect of TCGAN on gear fault diagnosis, we conduct different experiments. Model 1 is composed of real samples and simulates the fault diagnosis of imbalanced data sets in the case of small samples. Model 2 is composed of real samples and TCGAN-generated samples to generate samples instead of real samples to construct a balanced data set and compare the CNN recognition of the two models' rates.

The experimental results are shown in Table 6. It can be seen from the table that when the sample expansion is not performed, the CNN recognition rate is 82.5% when performing fault diagnosis on the unbalanced data set, and when the TCGAN sample expansion is performed, the imbalance of the data disappears. The CNN recognition rate is increased to 97.5%.

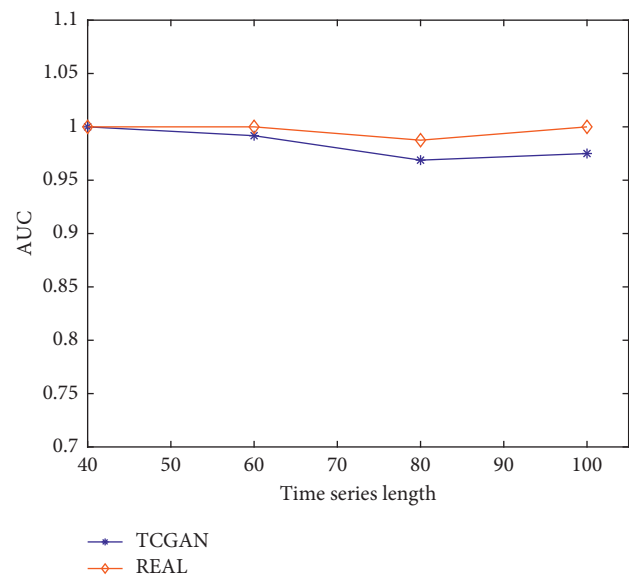


FIGURE 3: AUROC curve under training signal number 40.

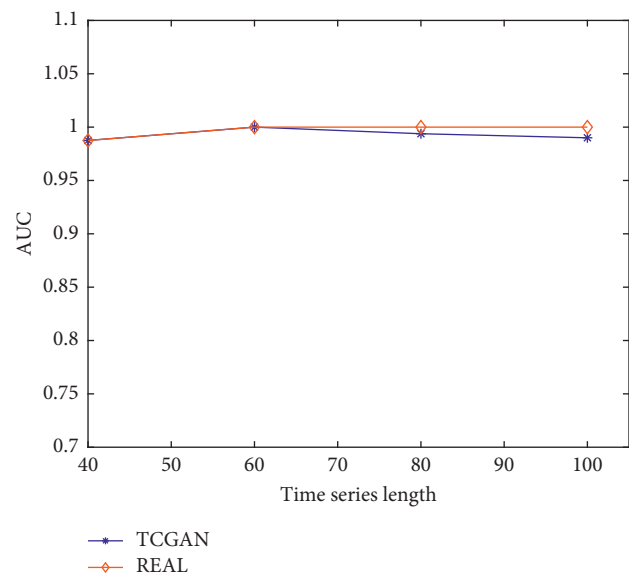


FIGURE 4: AUROC curve under training signal number 60.

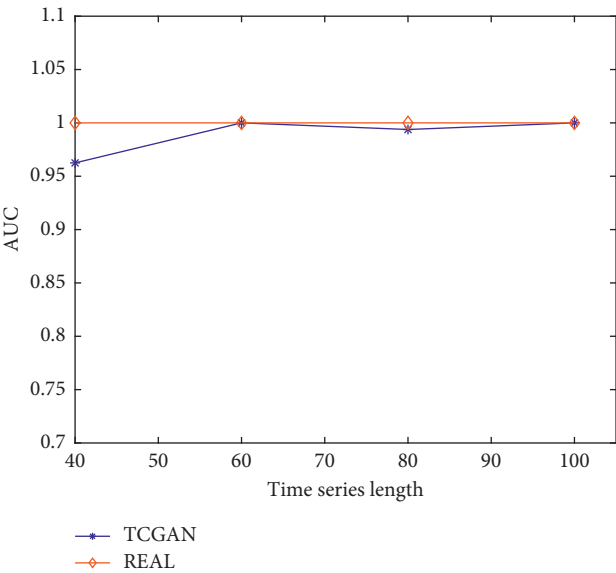


FIGURE 5: AUROC curve under training signal number 80.

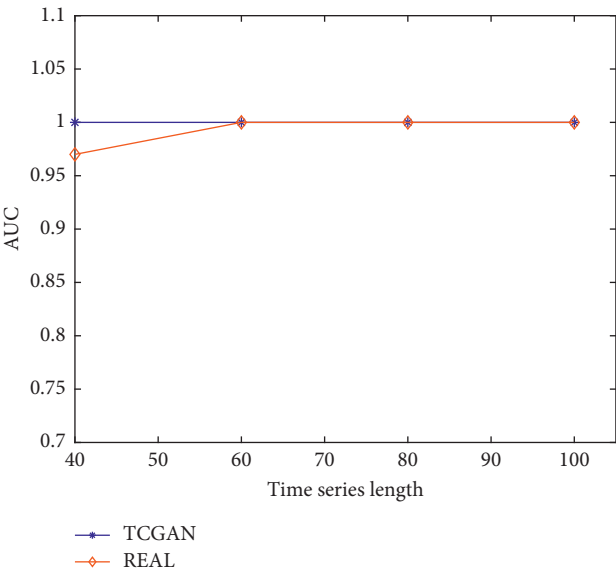


FIGURE 6: AUROC curve under training signal number 100.

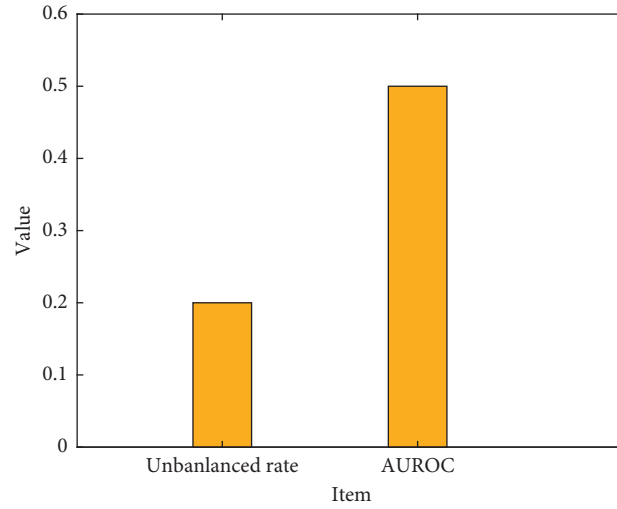


FIGURE 7: Unbalance rate and AROUC value before data enhancement.

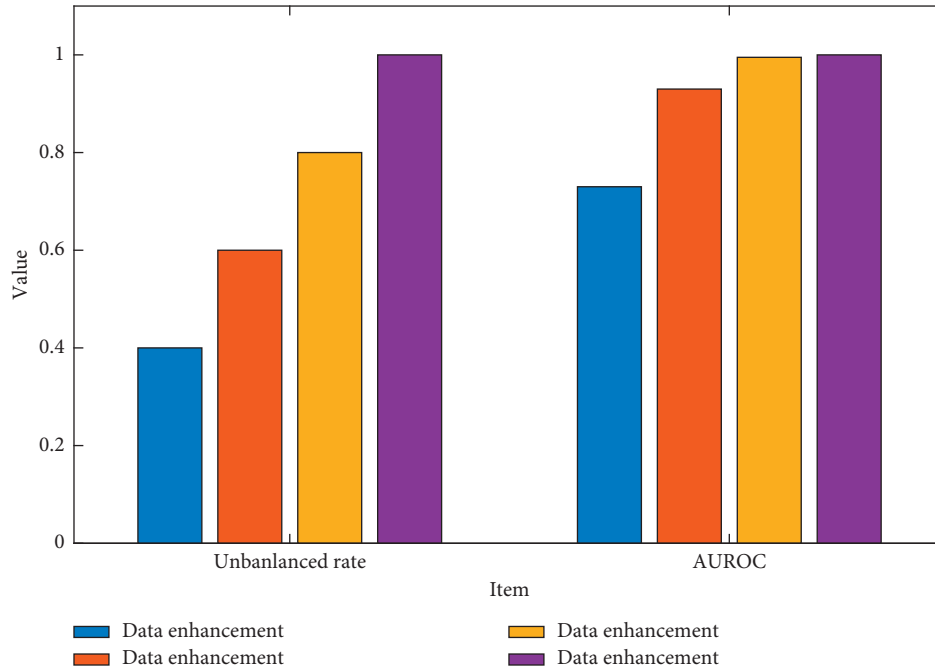


FIGURE 8: Unbalance rate and AROUC value after data enhancement.

TABLE 2: Comparison of characteristics of real signal and generated signal.

Feature	True signal	Generated signal
Root mean square	8.1995	8.0637
Square root amplitude	4.1837	3.9922
Absolute average	5.3995	5.211
Peak value	35.5857	36.2089
Peak-to-peak value	65.5838	66.6553
Waveform indicators	1.5178	1.5474
Peak shape indicator	3.6603	3.7759
Impulse indicator	5.5558	5.8427
Margin indicator	7.1703	7.6265
Skewness indicator	-0.2674	-0.4931
Kurtosis index	7.7696	8.5975
Amplitude spectral entropy	5.1932	5.1608
Envelope spectral entropy	5.2706	5.3431

TABLE 3: Bearing signal CNN experimental comparison.

Model	CNN training sample	CNN test sample
Model 1	100 groups of normal bearing signals and 30 groups of three types of faulty bearing signals	100 groups of normal bearing signals and 100 groups of three types of faulty bearing signals each
Model 2	100 groups of normal bearing signals, 30 groups of three types faulty bearing signals, and 70 groups of three types bearing signals	100 groups of normal bearing signals and 100 groups of three types of faulty bearing signals each

TABLE 4: CNN recognition rate of laboratory bearing data.

Quantity value	Model 1	Model 2
CNN recognition rate	65.5%	87.2%

TABLE 5: Comparison of real signal and generated signal of the cracked gear.

Feature	True signal	Generated signal
Root mean square	16.6302	12.6789
Square root amplitude	10.5792	8.4973
Absolute average	12.7646	10.0438
Peak value	50.1458	38.7464
Peak-to-peak value	94.5800	74.4775
Waveform indicators	1.3028	1.2624
Peak shape indicator	2.6719	2.6603
Impulse indicator	3.4811	3.3582
Margin indicator	4.2002	3.9694
Skewness indicator	-0.1508	-0.2328
Kurtosis index	3.5299	3.0871
Amplitude spectral entropy	4.2115	5.1942
Envelope spectral entropy	5.0374	5.1153

TABLE 6: CNN recognition rate of laboratory gear data.

Quantity value	Model 1	Model 2
CNN recognition rate	82.5%	97.5%

5. Conclusion

With the development of intelligent equipment, the data obtained based on the monitoring of mechanical equipment often have the characteristics of massive high dimensionality, nonstationarity, and nonlinearity. This makes the extracted initial fault characteristics unable to effectively identify the state and fault diagnosis of the mechanical equipment. In order to obtain high fault diagnosis accuracy, it is necessary to adopt effective dimensionality reduction technology to perform certain processing on the original high-dimensional fault feature set based on multidomain construction, remove redundant and irrelevant features, and extract low-dimensionality features that reflect the operating state of the equipment. Aiming at the problem of missing or incomplete category labels of fault samples, this paper studies the dimensionality reduction method of fault data sets based on semisupervised learning and proposes a new fault identification model based on semisupervised dimensionality reduction. The paper studies the fault diagnosis of rotating machinery in the case of small samples and unbalanced data. It designs a specific generative counter-measure network structure for the characteristic signals of

rotating machinery and applies it to rotating machinery data expansion and fault diagnosis field. We input the obtained dimensionality reduction result into the KNN classifier for training and learning and then realize the fault type recognition. The proposed method was verified with a data set of fault characteristics of a double-span rotor experimental rig. The results show that compared with other methods in the experiment, the KSELF method has stable dimensionality reduction ability and can obtain better dimensionality reduction effects and improve the classification accuracy rate. In future, we aim to enhance our work to achieve better results by using small training data sets. We aim to work on the proposed system so that it is able to diagnose a number of other types of machinery other than rotating machinery.

Data Availability

The data sets used are available from the corresponding author on reasonable request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This project was supported by the National Natural Science Foundation of China, under Project No. 61562009 (Research on Dynamic Optimization Problem Model and Algorithms Based on Enhanced Learning).

References

- [1] J. Lin and L. Qu, "Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis," *Journal of Sound and Vibration*, vol. 234, no. 1, pp. 135–148, 2000.
- [2] R. A. Collacott, *Mechanical Fault Diagnosis and Condition Monitoring*, Wiley distributed in the U.S.A. by Halsted Press, Hoboken, NJ, USA, 1977.
- [3] J.-H. Zhou, C. K. Pang, Z.-W. Zhong, and F. L. Lewis, "Tool wear monitoring using acoustic emissions by dominant-feature identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 2, pp. 547–559, 2011.
- [4] H. Y. Pan, "Research and Application on Fault Diagnosis of Rotating Machine Based on Variable Predictive Model Based Class Discriminate," Master, type, Thesis, subject, Mechanical engineering, HUNAN UNIVERSITY, Changsha, China, 2014.
- [5] S. Wang, J. Xiang, Y. Zhong, and Y. Zhou, "Convolutional neural network-based hidden Markov models for rolling element bearing fault identification," *Knowledge-Based Systems*, vol. 144, no. 15, pp. 65–76, 2018.
- [6] W. Jiang and X. Luo, "Research on unsupervised coloring method of Chinese painting based on an improved generative adversarial network[J]," *World Scientific Research Journal*, vol. 5, no. 11, pp. 168–176, 2019.
- [7] Z. Yang and L. Gao, *Wavelet Analysis and Fault Feature Extraction of Rolling bearing*, Springer, New York, NY, USA, 2014.
- [8] Z. Haiyang, W. Jindong, H. Hui, and G. Yiqi, "A feature extraction method based on HLMD and MFE for bearing clearance fault of reciprocating compressor," *Measurement*, vol. 89, pp. 34–43, 2019.
- [9] L. I. Hai and E. Huanan, "Fault diagnosis and treatment of rotor misalignment in wind turbines," *Electric Engineering*, vol. 1, 2018.
- [10] M. Hunter, A. Kourtellis, C. Ziomek, and W. Mikhael, "Fundamentals of modern spectral analysis," *IEEE Instrumentation and Measurement Magazine*, vol. 14, no. 4, pp. 12–16, 2011.
- [11] M. N. Dolmaz, Z. M. Hisarli, T. Usta♦mer, and N. Orbay, "Curie point depths based on spectrum analysis of aeromagnetic data, west anatolian extensional province, Turkey," *Pure and Applied Geophysics PAGEOPH*, vol. 162, no. 3, pp. 571–590, 2005.
- [12] P. D. Mcfadden, "Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration," *Journal of Vibration and Acoustics*, vol. 108, no. 2, pp. 165–170, 1986.
- [13] X. M. Li, Y. Lv, C. C. Yi, and Z. X. HUANG, "Research of mechanical fault feature extraction method based on tensor singular spectrum decomposition," *Machinery Design & Manufacture*, vol. 19, no. 7, pp. 56–59, 2017.
- [14] H. W. Wang, J. Cui, F. Xiao, and Z. DONG, "fault diagnosis of rolling bearings based on dynamics and Zoom spectrum," *Journal of Beijing University of Technology*, vol. 40, no. 10, pp. 1447–1453, 2014.
- [15] Y. Li, M. Xu, Y. Wei, and W. Huang, "A New Rolling Bearing Fault Diagnosis Method Based on Multiscale Permutation Entropy and Improved Support Vector Machine Based Binary tree," *Measurement*, vol. 77, pp. 80–94, 2016.
- [16] L. H. Wang, X. P. Zhao, J. X. Wu, and Y. Y Xie, "Motor Fault Diagnosis Based on Short-Time Fourier Transform and Convolutional Neural Network," *Chinese Journal of Mechanical Engineering*, vol. 30, no. 6, pp. 1357–1368, 2017.
- [17] H. Xu and G. Chen, "An intelligent fault identification method of rolling bearings based on LSSVM optimized by improved PSO," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 167–175, 2013.
- [18] W. Hu and B. C. Si, "Soil water prediction based on its scale-specific control using multivariate empirical mode decomposition," *Geoderma*, vol. 193-194, pp. 180–188, 2013.
- [19] Y. Li and N. Q. Shu, "transformer fault diagnosis based on fuzzy clustering and complete binary tree support vector machine," *Transactions of China Electrotechnical Society*, vol. 31, no. 4, pp. 64–70, 2016.
- [20] R. Zhang, L. Wu, X. Fu, and B. Yao, "Classification of bearing data based on deep belief networks," in *Proceedings of the 2016 Prognostics and System Health Management Conference* Chengdu, China, IEEE, 2017.
- [21] I. Yani, Y. Resti, and F. RestiBurlian, "Identification of bearing failure using signal vibrations," *Journal of Physics: Conference Series. IOP Publishing*, vol. 1007, no. 1, p. 012067, 2018.
- [22] D.-T. Hoang and H.-J. Kang, "Convolutional neural network based bearing fault diagnosis," *Intelligent Computing Theories and Application*, pp. 105–111, 2017.
- [23] H. C. Wu, W. L. Fan, C. S. Tsai, and J. C. Ying, "An Image Authentication and Recovery System Based on Discrete Wavelet Transform and Convolutional Neural networks," *Multimedia Tools and Applications*, vol. 5, pp. 1–25, 2021.
- [24] P. Luo, Y. Yang, and J. S. Cheng, "Adaptive Fisher-based deep convolutional neural network method," *Journal of Electronic Measurement and Instrumentation and its application in gear fault diagnosis*, vol. 32, no. 1, pp. 33–39, 2018.
- [25] Y. J. Jiang and L. Wang, "Cognitive network end to end situational evaluation algorithm based on BP-DBN," *Computer Technology and Development*, vol. 24, no. 11, pp. 148–151, 2014.
- [26] J. Pei, K. Zhong, J. Li, J. Xu, and X. Wang, "ECNN: evaluating a cluster-neural network model for city innovation capability," *Neural Computing & Applications*, pp. 1–13, 2021.