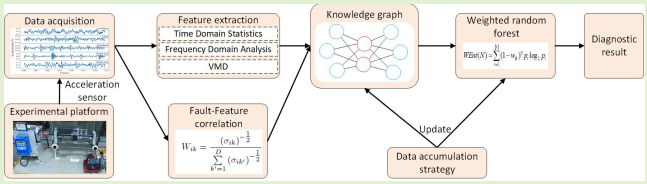
Fault Diagnosis of Rolling Bearing Based on Knowledge Graph With Data Accumulation Strategy

Xiangqu Xiao, Chaoshun Li, *Member, IEEE*, Jie Huang, and Tian Yu

***Abstract*—The fault diagnosis of rolling bearing plays an important role in ensuring the safe and stable operation and maintenance of rotating machinery. Traditional bearing fault diagnosis methods fail to consider the correlation between faults and characteristics and do not take full advantage of the ever-increasing monitoring data. Thus, a bearing fault diagnosis framework based on knowledge graph (KG) and data accumulation strategy is proposed. First, the entities of the KG are defined based on multiple features extracted from the time domain, frequency domain, and time–frequency domain of bearing vibration data collected by the vibration sensors. Then, the feature-fault correlation as the edges is designed and calculated to establish a KG framework together with the entities. In addition, a weighted random forest algorithm is proposed as a reasoning algorithm for the KG, making full use of the feature-fault correlation to improve the accuracy of bearing fault classification. Finally, a data accumulation strategy is designed to continuously increase the size of the training dataset of KG. Relevant parameters are updated in the process to make the results produced by the inference algorithm more accurate. The advantage of the proposed method was demonstrated by a comparison with several models for the same circumstance. The test results showed that the proposed method was promising and it had good prediction accuracy and robustness for different working conditions.**



***Index Terms*— Data accumulation strategy, fault diagnosis, feature-fault correlation, knowledge graph (KG), weighted random forest (WRF).**

I. Introduction

ROTATING machinery plays an important role in industrial systems. When they fail, they tend to cause large economic losses and even major accidents in severe cases [1]. Rolling bearings are an important part of rotating machinery, and it is also a relatively fragile part. According to statistics, bearing failures account for about 30% of rotating machinery [2]. Therefore, it is of great significance to carry out fault diagnosis research for rolling bearings, realize rapid and accurate classification of fault categories, and ensure stable and efficient operation of rotating machinery.

Manuscript received 29 July 2022; accepted 14 August 2022. Date of publication 31 August 2022; date of current version 30 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 51879111, in part by the Applied Fundamental Frontier Project of Wuhan Science and Technology Bureau under Grant 2018010401011269, and in part by the Hubei Provincial Natural Science Foundation of China under Grant 2019CFA068. The associate editor coordinating the review of this article and approving it for publication was Dr. Dong Wang. *(Corresponding author: Chaoshun Li.)*

The authors are with the School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: d201981042@hust.edu.cn; csli@hust.edu.cn; hjie@hust.edu.cn; d201880943@hust.edu.cn).

Digital Object Identifier 10.1109/JSEN.2022.3201839

In the early stage of industrial development, bearing fault detection mainly relied on the experience of technicians. With the emergence and development of computer technology, a spectrum analysis has changed from theory to reality. Bearing characteristics and fast Fourier transform technology can be used to determine bearing faults. The subsequent resonance demodulation technology has also been widely used in the field of bearing fault diagnosis. With the rapid development of computer technology, machine learning and deep learning have gradually become the mainstream technology for bearing fault diagnosis.

The fault diagnosis process for rolling bearing based on machine learning can usually be divided into two steps. First, time-domain statistical analysis and other transform-domain analysis methods are used to extract the typical characteristics of the original signal, and then, the classifier is designed to classify and diagnose the fault type. Among them, the Bayes criterion, linear discriminant function, and nonlinear discriminant function are often used to design pattern classifiers. Common models include support vector machine (SVM) [3], [4], [5], K-nearest neighbor (KNN) [6], [7], naive Bayes (NB) [8], and artificial neural network (ANN) [9], [10]. This type of method is usually simple to implement, but it may show a deficiency in the face of complex fault conditions and has certain limitations. The deep-learning technology

developed in recent years exhibits the powerful ability to unify the traditional two-step task within a whole process. Deep learning [11], [12], [13] can establish a complex nonlinear relationship between input and output through a multilayer network structure. Through complex nonlinear transformations, the underlying features can be automatically learned from the training data and gradually formed through the multilayer network. The abstract high-level representation overcomes the shortcomings of manually selected features, adaptively obtains the optimal feature combination, realizes end-to-end learning, and completes the fault diagnosis of the bearing. These methods based on deep learning have achieved good results in the effect of bearing fault diagnosis. However, they seldom consider how to handle the ever-increasing monitoring data of the fault bearing. The continuous accumulation and use of the monitoring data would augment the fault knowledge and thus effectively improve the accuracy of the diagnosis method. Therefore, it is attractive to study the knowledge graph (KG) for fault diagnosis of bearing considering the fact that the data will increase and the fault knowledge is not static in real industrial applications.

The KG is a semantic network proposed by Google in 2012, which stores knowledge in the form of a network [14]. The KG expresses facts in the form of entities and relationships. The entities are the nodes in the graph and the relations are the edges connecting the nodes in the graph [15]. KG can effectively solve the problems of automatic question and answer, personalized recommendation, and intelligent information retrieval [16], [17], and it is widely used in antifinancial fraud and diagnosis and treatment decision-making [18]. According to the coverage of the KG, it can be divided into general KG and industry KG. General KGs are oriented to multiple fields or even the whole field. The emphasis is on the breadth of knowledge and the number of entities and relationships involved is often massive. The representative general KGs are WordNet [19], DBpedia [20], wikidata [21], and so on. Industry KGs are also called domain KGs. In this article, the domain KG of bearing fault diagnosis is studied. This type of KG is constructed using industry data for a specific field and emphasizes the professionalism and reliability of knowledge, so as to provide relatively accurate and detailed information for personnel in the industry, for example, the financial KG [22] and the medical KG [23]. However, in the field of rolling bearing fault diagnosis, there is no more representative research.

So far, the professional KG is blank in the field of bearing fault diagnosis, and the related methods and technologies are still in their infancy. Refer to the system of KG in other fields to get the corresponding key steps: extraction of vibration signal features, quantification of the correlation between features and faults, the reasoning of fault categories based on the constructed KG, and knowledge accumulation as the number of data increases. Vibration signal feature extraction can obtain information that highlights the differences between different fault categories and weaken or remove the information that has nothing to do with fault diagnosis. The vibration signal feature extraction methods can be roughly divided into time-domain analysis, frequency-domain analysis, and time–frequency analysis. These three methods will be used to extract the characteristics of vibration signals to make up for the shortcomings of a single method, thereby improving the accuracy of fault classification. In addition, the feature weighting coefficient in soft clustering is used to quantify the correlation between features and faults, an improved random forest algorithm is proposed for fault category inference, the correlation between features and faults is used to maximize faults classification accuracy, and a data accumulation strategy is proposed.

The main scientific contributions of this research can be listed as follows.

1) A feature weighting coefficient is defined, which can quantify the correlation between faults and features, indicate the importance of each feature in different fault categories, and effectively reduce the impact of invalid features on the accuracy of fault diagnosis.
2) An innovative bearing fault diagnosis model based on KG is proposed, and an improved random forest algorithm is proposed as the reasoning algorithm of KG to realize fault diagnosis. In this algorithm, the split nodes are divided by feature importance, and the correlation between features and faults is used to effectively improve the classification accuracy.
3) A strategy for the accumulation of KG data is proposed, which can realize the effective accumulation of data and the update of model parameters, maximize the use of data, and achieve the improvement of fault classification accuracy.

The rest of this article is organized as follows. Section II introduces the random forest algorithm and the structure of the fault KG. Section III details the construction method of the KG, random forest improvement strategy, and KG update and data accumulation strategy. The experimental process and results are presented in Section IV. Finally, the conclusion is given in Section V.

II. Related Works

A. Knowledge Graph

The KG is a structured semantic knowledge base, composed of a series of nodes, edges, and attributes, and uses triples to describe the data. The form of triples is generally "entity–relation–entity." A node is a manifestation of an entity (or concept) in the physical world. The entity is the actual thing, and the concept is usually a collection of things. Edge refers to the relationship between nodes, and the relationship is the various connections between entities.

There are a series of processes and operations in the construction and application of KG. The following parts are focused on entity acquisition, relation acquisition, knowledge reasoning, and cumulative update.

1) *Entity Acquisition:* For the KG in the field of bearing vibration, the entity mainly includes the characteristics of the vibration signal and the typical state of the bearing. The characteristics of the vibration signal can be obtained through the time-domain feature function, frequency-domain feature function, and variational mode decomposition (VMD) [24].
2) *Relationship Acquisition:* For the relationship between entities, specific data need to be used to describe the

strength of the correlation between nodes. In the bearing fault KG, the relationship between features and faults is often very important, which can be obtained through statistical methods or other methods.

3) *Knowledge Reasoning:* Knowledge reasoning is an important part of the KG, and other unknown things can be derived from known knowledge reasoning. Knowledge reasoning methods can be divided into two categories: logic-based reasoning [25], [26] and graph-based reasoning [27]. In terms of realization, reasoning can be realized through logic rules and logic symbols and also through machine learning algorithms and neural network models.
4) *Accumulate and Update:* The accumulation and updating of knowledge are one of the advantages of KG. The amount of information and knowledge will continue to increase over time, so the content of the KG must also keep pace with the times. Constantly increasing knowledge and data will make the function of the KG more powerful.

B. Random Forest

It is a machine learning method composed of multiple decision tree models [28]. Randomization is used in the selection of samples and features to generate multiple decision trees and then summarize the results of all decision trees to obtain the final classification result.

The traditional random forest classification algorithm mainly has four essential contents.

1) *Randomly Select Samples:* Sampling with replacement from the original training sample set of N number to obtain N samples to form a new training set.
2) *Randomly Select Features:* In the process of building a decision tree, when a node splits, the algorithm needs to extract k_0 features from the k features without replacement.
3) *Build a Decision Tree:* Select the optimal split feature attribute of the current node according to the information gain, gain ratio, or Gini index of k_0 features, and repeat the process until the construction is completed.
4) *Random Forest Voting Mechanism:* When a random forest is used as a classification model, the result that appears most in the decision tree is selected as the final classification result.

Random forest classification algorithm has the advantages of not being easy to produce overfitting, low feature requirements, fewer adjustment parameters, and so on and has a wide range of application prospects in the field of machine learning. In this article, we will try to improve the most characteristic selection method of decision tree node splitting in a random forest so that it can make full use of the importance of each feature under different categories to improve the final classification result.

III. Proposed Method

The relevant content of the bearing fault knowledge graph (BFKG) will be introduced in this section. Aiming at the different importance of each feature of different faults, a fault-feature correlation calculation method is proposed to establish the "edge" between nodes of the KG. At the same time, an improved random forest algorithm is proposed, which aims to use the fault-feature correlation to assist in selecting the optimal feature when the decision tree node is split. Finally, a data accumulation strategy is proposed, the highly reliable results after each test are selected and added to the next training set, and the training set data are continuously increased to improve the accuracy of bearing fault classification. The overall process is shown in Fig. 1.

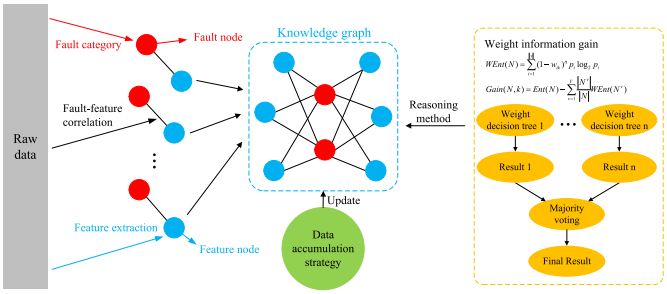


Fig. 1. General pipeline of mechanical fault diagnosis with the proposed method.

TABLE I
FEATURE PARAMETERS OF TIME DOMAIN

Time-domian feature parameters	
$p_1 = \frac{\sum_{n=1}^{N} x(n)}{N}$	$p_7 = \frac{\sum_{n=1}^{N} (x(n)-p_1)^4}{(N-1)p_2^4}$
$p_2 = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-p_1)^2}{N-1}}$	$p_8 = \frac{p_5}{p_4}$
$p_3 = \left(\frac{\sum_{n=1}^{N} \sqrt{\lvert x(n)\rvert}}{N}\right)^2$	$p_9 = \frac{p_5}{p_3}$
$p_4 = \sqrt{\frac{\sum_{n=1}^{N} (x(n))^2}{N}}$	$p_{10} = \frac{p_4}{\frac{1}{N}\sum_{n=1}^{N} \lvert x(n)\rvert}$
$p_5 = \max \lvert x(n)\rvert$	$p_{11} = \frac{p_5}{\frac{1}{N}\sum_{n=1}^{N} \lvert x(n)\rvert}$
$p_6 = \frac{\sum_{n=1}^{N} (x(n)-p_1)^3}{(N-1)p_2^3}$	

Where $x(n)$ is a signal series for n = 1,2,...,N, N is the number of data points.

A. Feature Extraction

A sliding window with a length of 2400 is used to divide the measured signal into smaller signal segments. Due to the high sampling frequency, the length of 2400 already contains the necessary health information for bearing. Feature extraction of signal segments is very important for fault diagnosis. In this article, signal features were extracted from three aspects: time domain, frequency domain, and time–frequency domain, so as to obtain more complete signal features.

The multiple features of the signal can effectively improve the accuracy of bearing fault diagnosis. This article references [29], [30] to extract relevant features. The time-domain features include 11 indicators such as signal mean value and kurtosis, as shown in Table I. Some of the features are used to reflect the amplitude and energy of the signal in the

TABLE II
FEATURE PARAMETERS OF FREQUENCY DOMAIN

Frequency-domian feature parameters	
$p_{12}=\frac{\sum_{l=1}^{L}s(l)}{L}$	$p_{18}=\sqrt{\frac{\sum_{l=1}^{L}(f_l-p_{16})^2 s(l)}{L}}$
$p_{13}=\frac{\sum_{l=1}^{L}(s(l)-p_{12})^2}{L-1}$	$p_{19}=\sqrt{\frac{\sum_{l=1}^{L}f_l^4 s(l)}{\sum_{l=1}^{L}f_l^2 s(l)}}$
$p_{14}=\frac{\sum_{l=1}^{L}(s(l)-p_{12})^3}{L(\sqrt{p_{13}})^3}$	$p_{20}=\frac{\sum_{l=1}^{L}f_l^2 s(l)}{\sqrt{\sum_{l=1}^{L}s(l)\sum_{l=1}^{L}f_l^4 s(l)}}$
$p_{15}=\frac{\sum_{l=1}^{L}(s(l)-p_{12})^4}{Lp_{13}^2}$	$p_{21}=\frac{p_{18}}{p_{16}}$
$p_{16}=\frac{\sum_{l=1}^{L}f_l s(l)}{\sum_{l=1}^{L}s(l)}$	$p_{22}=\frac{\sum_{l=1}^{L}(f_l-p_{16})^3 s(l)}{Lp_{18}^3}$
$p_{17}=\sqrt{\frac{\sum_{l=1}^{L}f_l^2 s(l)}{\sum_{l=1}^{L}s(l)}}$	$p_{23}=\frac{\sum_{l=1}^{L}(f_l-p_{16})^4 s(l)}{Lp_{18}^4}$

Where $s(l)$ is a spectrum for l = 1,2,...,L, L is the number of spectrum lines; f_l is the frequency value of the l-th spectrum line.

time domain, and the other part of the feature reflects the distribution of the signal in the time series. Frequency-domain features are mainly a series of statistical indicators extracted in the frequency domain after the Fourier transform of the signal. In this article, a total of 12 frequency-domain features have been extracted, as shown in Table II. Some of the features are used to describe the convergence of the signal spectrum, and the other part of the features reflects the distribution of main frequencies in the signal frequency components.

The signal time–frequency domain characteristics involved in this article mainly include energy characteristics and singular value characteristics. In this article, the VMD method is used to transform the signal to obtain a series of modes, and then, the energy characteristics and singular value characteristics of these modes are obtained. The original signal is decomposed by VMD to obtain a series of modes $C=\{C_1, C_2, \ldots, C_M\}$, where M is the number of modes after decomposition—in this article, M is set to 4. $E_1, E_2, \ldots, E_m$ is the energy value corresponding to each mode. The specific calculation equation is shown in the following equation:

$$E_m=\int_{-\infty}^{+\infty}|C_m|^2 dt. \tag{1}$$

The decomposed modes of the signal form a matrix $A=[C_1, C_2, \ldots, C_M]^T$. The singular value decomposition of the matrix can extract the singular value characteristics of the matrix.

B. Fault-Feature Correlation

For different bearing faults (i.e., different working conditions), the same characteristics usually have different importance. In addition, the features extracted from the vibration signal may have redundancy or noise. Direct use of these features may reduce the accuracy of classification. At the same time, it cannot reflect the contribution of each feature to different fault types. In this article, a method for calculating the correlation between faults and features is presented. The specific equations are given as follows:

$$w_{ik}=\frac{(\sigma_{ik})^{-\frac{1}{2}}}{\sum_{k'=1}^{D}(\sigma_{ik'})^{-\frac{1}{2}}};\quad i=1,2,\ldots,C;\ k=1,2,\ldots,D \tag{2}$$

$$\sigma_{ik}=\sum_{j=1}^{N}u_{ij}(x_{jk}-v_{ik})^2 \tag{3}$$

$$v_{ik}=\frac{\sum_{j=1}^{N}u_{ij}x_{jk}}{\sum_{j=1}^{N}u_{ij}}. \tag{4}$$

In (2), w_{ik} represents the correlation between the kth features and the ith fault, D is the total number of the feature categories, C is the number of fault categories, σ_{ik} represents the sum of the distances from the kth dimension data of all signal features belonging to the ith category to the corresponding data center, and (3) is the calculation equation of σ_{ik}. In (3), u is the membership matrix, where u_{ij} represents the probability that the jth data belong to the ith type of fault, x_{jk} represents the kth feature dimension data of the jth sample, v_{ik} represents the kth dimension data center belonging to the ith fault data, and (4) is the calculation equation of v_{ik}. The symbols in (4) have been introduced before.

It can be seen from (2) that for the kth dimension feature of the data belonging to the ith fault type, the closer to the data center of this dimension, the larger w_{ik}, indicating that the kth dimension of the data under the ith fault type is more concentrated; then, the greater the correlation between this feature and the fault type. At the same time, it can be found that the sum of the correlations between all features and the same fault is 1. Using this method, the correlation of each feature to each fault type can be obtained, which provides a theoretical basis for the establishment of the "edge" of the KG.

C. Weighted Random Forest Algorithm

For the bearing fault KG, the knowledge reasoning process is very important. After the KG is established, the classification of bearing faults needs to be realized through the inference process. The random forest algorithm is used as the reasoning method of the KG. Information entropy is a commonly used index to measure the purity of a sample set, and it is usually used for optimal feature selection when dividing nodes in a decision tree. The traditional information entropy does not use the correlation of each feature to the fault, that is, does not use the "edge" information in the KG, so the traditional random forest is not suitable to be used here as an inference algorithm to classify fault types. In this article, a weighted random forest (WRF) algorithm is proposed for this point. Adjust the traditional information entropy equation, aiming to use the fault-feature correlation to assist the decision tree node split. The formula is given by

$$W\mathrm{Ent}(N)=\sum_{i=1}^{|y|}(1-w_{ik})^{\eta}p_i\log_2 p_i \tag{5}$$

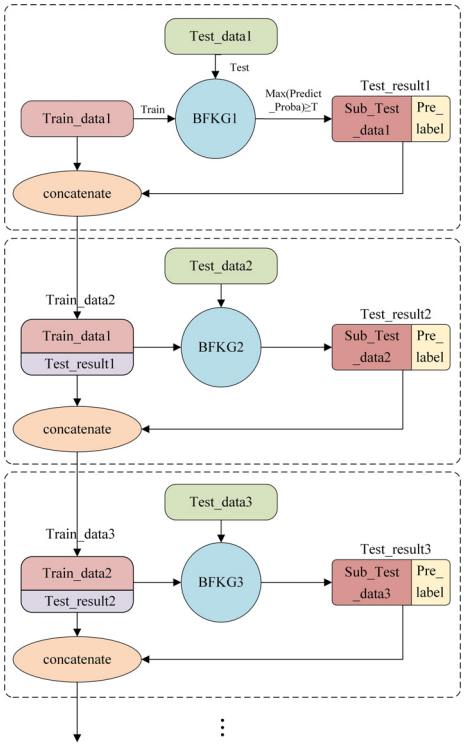


Fig. 2. Flowchart of data accumulation. BFKG is the diagnostic method proposed in this article. If the maximum predicted probability of the data in Test_data is greater than the given threshold value T, the data are recorded in Sub_Test data, Pre_label is the prediction label of the data, and Sub_Test data and Pre_label form Test_result. Train_data and Test_result of the previous model are concatenated to form Train_data of the next model.

where w_{ik} represents the correlation between the kth features and the ith fault, $|y|$ represents the number of current branch categories, p_i represents the proportion of the ith data in the set N, and η is the adjustment parameter—in this article, η is set to 1/3.

WRF uses the information gain ratio splitting criterion, where the information gain ratio equation obtained by dividing the sample N using the feature k is given by

$$\text{Gain}(N,k) = \text{Ent}(N) - \sum_{v=1}^{V} \frac{|N^v|}{|N|} W\text{Ent}(N^v) \tag{6}$$

where $\text{Ent}(\cdot)$ is the traditional information entropy equation; V is the number of branches produced by the split, in this article, V is set to 2; and $|N^v|$ is the number of samples on the vth branch.

D. KG Data Accumulation Strategy

For data-driven models, the scale of data is very important. The larger the dataset used for training, the better the effect of the test results. Therefore, it is meaningful to increase the data scale and increase the degree of data utilization.

In this article, a data accumulation strategy designed for the KG is proposed, as shown in Fig. 2. This strategy selects the appropriate test data and the corresponding prediction results as the new data for the next model training after each data test. The main work of this strategy is to introduce a threshold T. The function of this threshold is to screen suitable data from the test results, that is, select the data with the maximum prediction probability not less than T in the results, and the corresponding label value is the label value predicted by the model.

This threshold is often related to the performance of the model on the dataset. When the model has a good classification effect on the dataset, the threshold T is often small so that more test results can be screened out for the next training. When the classification effect of the model on the dataset is poor, the threshold often needs to be larger so that there are a few error samples as possible in the filtered results.

The specific implementation steps are given as follows.

Step 1: Divide the original dataset into multiple datasets, one of which is used as the training set and the rest as the test set.

Step 2: Determine the membership matrix u according to the training sample data. When the jth sample data belong to the ith type of fault, $u_{ij} = 1$; otherwise, $u_{ij} = 0$. Calculate the sample center matrix v of each type of data according to (4). Calculate the correlation matrix W between features and faults according to (2) and (3).

Step 3: Train the BFKG model and get the predicted label of the test set sample and the probability matrix.

Step 4: Take a test sample for analysis. If the maximum value of the probability matrix of this test sample is not less than T, add this test sample and the predicted label combination to the next training sample, add the probability matrix to the previous degree of membership as the membership matrix of the new sample in the matrix, and prepare for the next training.

IV. Experimental Results and Comparisons

Three bearing datasets will be used, two of which are provided by the Mechanical Failure Prevent Technology Association and Case Western Reserve University (CWRU). The other is collected by the mechanical fault comprehensive simulation experiment system in our laboratory, including various states of the bearing. There cases are designed on these datasets. Case I is a routine experiment, which compares the classification accuracy of BFKG with other methods to verify the advantages of BFKG in diagnosis. In Case II, after training the model with the original training set, the data worth keeping in each test set will add to the training set for the next model training. By comparing with the situation of not adding data to the training set, the effectiveness of the data accumulation strategy in the data growth mode is verified. By comparing with other classification methods using data accumulation strategies, the advantage of using data accumulation in KG is verified. In Case III, experiments are conducted by constructing an imbalanced dataset on the CWRU dataset, which verifies the strong robustness of the proposed method.

A. Datasets Descriptions

1) MFPT Datasets (MT0 and MT1): The MFPT dataset is provided by the Mechanical Failure Prevent Technology Association [31]. This dataset contains data from the bearing test

TABLE III
TOTAL 28 WORKING CONDITIONS OF THE CWRU DATASET

Health condition	0W	735.5W	1471W	2206.5W
Normal	√	√	√	√
OR 0.1778mm	√	√	√	√
OR 0.5334mm	√	√	√	√
IR 0.1778mm	√	√	√	√
IR 0.5334mm	√	√	√	√
B 0.1778mm	√	√	√	√
B 0.5334mm	√	√	√	√

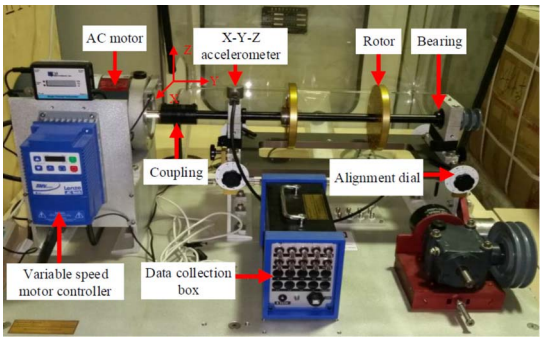


Fig. 3. Laboratory bearing experimental platform.

rig and three real-world failure data. The data from the bearing test rig include three kinds of baseline datasets, ten kinds of outer race faults datasets, and seven kinds of inner race faults datasets. The signal data of 15 states were selected for the experiment. The specific components include baseline condition data; outer ring failure data under seven different load conditions, with loads of 11.34, 22.68, 45.36, 68.04, 90.72, 113.40, and 136.08 kg; and inner ring fault data for seven different load conditions, loads of 0, 22.68, 45.36, 68.04, 90.72, 113.40, and 136.08 kg. The sampling frequency of bearing vibration signal is 48828 Hz, and the sampling time is 3 s. The number of samples in each state is 50, and the total number of samples in the dataset is 750.

2) CWRU Datasets (CU0 and CU1): The Electrical Engineering Laboratory of CWRU provides a bearing vibration dataset, which is widely used in the field of bearing fault diagnosis [32]. The testbed consisted of a 2-hp motor, torque transducer/encoder, and dynamometer. The dataset contains four bearing states, including normal (N), outer race fault (OR), inner race fault (IR), and ball fault (B). The dataset contained four different motor load conditions: 0, 735.5, 1471, and 2206.5 W. The signal data of 28 states were selected for the experiment. The specific components include normal state data under four different loads: outer ring fault, inner ring fault, and rolling element fault under eight different loads and fault diameters. Specific information can be found in Table III. The number of samples in each state is 50, and the total number of samples in the dataset is 1400. The signal sampling frequency was 12 kHz, and the sampling time was 10 s.

3) MFS Datasets (MS0 and MS1): In addition, more bearing data were produced from our own laboratory. Fig. 3 shows the experiment rig in our laboratory. The test rig is a mechanical fault comprehensive simulation experiment system (MFS),

Fig. 4. Bearings with different faults. (a) Outer race fault. (b) Inner race fault. (c) Ball fault. (d) Combination fault.

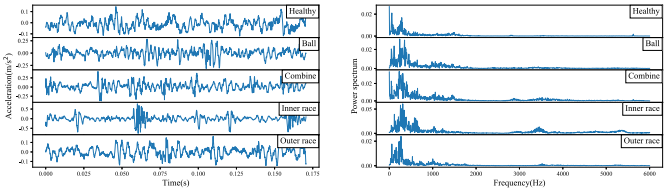


Fig. 5. Time- and frequency-domain outputs of samples in the MFS dataset.

TABLE IV
DATA DESCRIPTION OF CASE I

Dataset	Data description	Train	Test
MT0	15 classes of vibration data consists of baseline, Inner Race Fault and Out Race Fault with different loads.	15×40	15×10
CU0	28 classes of fans end vibration data. They consist of N, IRF, ORF, and BF with loads ranging from 0 to 3 loads and fault diameter are 0.1778 and 0.5334 mm.	28×40	28×10
MS0	15 classes of vibration data. They consist of N, IRF, ORF, BF, and CF, the speed ranging from 10rev/s to 30rev/s.	15×48	15×12

an experimental device produced by Spectra Quest in USA that can simulate bearing fault of rotating machinery. The system is composed of a mechanical fault comprehensive simulation experiment platform, an experiment platform kit, sensors, a data acquisition system, a PC terminal, and data acquisition and analysis software. The dataset of this experiment included normal data (N), outer race fault data (OR), inner race fault data (IR), ball fault data (B), and combined fault data (C). A detailed description of each fault condition is shown in Fig. 4. The speed of the shaft of the experiment rig is 10, 20, and 30 rev/s. The signal sampling frame was 12 kHz and the sampling time was 12 s. The signal data of 15 states were selected for the experiment. The specific components include outer ring fault, inner ring fault, rolling element fault, combined fault, and normal at three different rotational speeds. The specific information is shown in the Table IV. The number of samples in each state is 60, and the total number of samples in the dataset is 900. Time–frequency diagrams of samples with part of heath states are given in Fig. 5.

The datasets were processed differently according to different experiments. In Case I, each dataset is simply divided into a training set and a test set. The training set data account for 80% of the original data, and the rest 20% of the dataset is the test set; the new dataset obtained by processing the three original datasets is named MT0, CU0, and MS0, as shown in Table V. In Case II, each dataset is equally divided into five

TABLE V
DATA DESCRIPTION OF CASE II

Dataset	Data description	Train	Test
MT1	15 classes of vibration data consists of baseline, Inner Race Fault and Out Race Fault with different loads.	15×10	15×10×4
CU1	28 classes of fans end vibration data. They consist of N, IRF, ORF and BF with loads ranging from 0 to 3 loads and fault diameter are 0.1778 and 0.5334 mm.	28×10	28×10×4
MS1	15 classes of vibration data. They consist of N, IRF, ORF, BF, and CF, the speed ranging from 10rev/s to 30rev/s.	15×12	15×12×4

TABLE VI
TOTAL 15 WORKING CONDITIONS OF THE MFS DATASET

Health condition	10rev/s	20rev/s	30rev/s
N	√	√	√
OR	√	√	√
IR	√	√	√
B	√	√	√
C	√	√	√

parts. One of the datasets is used as the initial training set, and the remaining four are used as test sets to test the effect of the data accumulation strategy. There are three original datasets. The new datasets obtained after processing are named MT1, CU1, and MS1, as shown in Table VI.

B. Case I: Verification of Classification Effect of BFKG

In Case I, BFKG is compared with other classification models, including RF, SVM, ANN, 1-D convolutional neural network (1-D CNN), CNN-PCA-FCM [33], improved random forest [34], deep residual shrinkage networks [35], graph convolutional network (GCN) [36], and other WRFs [37]. The models compared include traditional machine learning algorithms and neural network models. The input of 1-D CNN and CNN-PCA-FCM is a signal segment with a length of 2400 because they can automatically extract deep features. It is verified that BFKG has certain advantages over other classification models in fault classification. The models are compared on datasets CU0, MT0, and MS0. The hyperparameter settings of SVM, RF, and other models are listed in Table VII.

The results of the ten methods in Case I on three datasets (CU0, MT0, and MS0) are shown in Table VIII, and the best results on each dataset are highlighted in bold. It can be seen from the table that BFKG shows the superiority of classification on multiple datasets. In order to show the superiority of BFKG more intuitively, the results are presented in the form of bar graphs in Fig. 6. The experiment shows that the use of feature-fault correlation for decision tree node splitting attribute selection can promote the classification of random forests, that is, the WRF reasoning algorithm based on the KG has certain advantages compared with other machine learning algorithms and simple neural network models. GCN utilizes the relationship between samples, and the obtained fault classification accuracy is better. Improved RF and WRF-OOB have been improved on traditional RF, and both have good results.

TABLE VII
SETTINGS OF OTHER CLASSIFICATION METHODS

Method	Structure description
SVM	C=1, Kernel: rbf, gamma=$1/f_n$, where f_n represents the dimension of the input feature vector, using rbf kernel.
RF	The number of decision trees in a random forest is 40, which is consistent with the number of decision trees in BFKG.
ANN	Number of neurons per layer:(number of feature)-256-512-256-(number of classes), Epochs=500, batch_size=128.
1D-CNN	Structure of each layer:Input-Conv1D-MaxPooling-Conv1D-MaxPooling-Flatten-Dense-Dropout-Dense, Epochs=50, batch_size=32.
CNN-PCA-FCM	The number of principal components extracted by PCA is 10, and the remaining parameters refer to the original paper.
Improved RF	Detailed parameters are set at [34].
DRSN-CS	Detailed parameters are set at [35].
GCN	Contains two graph convolutional layers and a fully connected layer.Number of neurons per layer:(number of feature)-32-32-(number of classes), Epochs=500.
WRF-OOB	The parameters are consistent with BFKG.

TABLE VIII
COMPARISON OF BFKG AND OTHER METHODS

Dataset	MT0	CU0	MS0
BFKG	**0.8500**	**0.9857**	**0.9556**
RF	0.8350	0.9750	0.9444
SVM	0.8400	0.9678	0.7944
ANN	0.7200	0.9607	0.9389
1D-CNN	0.7300	0.9607	0.8167
CNN-PCA-FCM	0.8350	0.9500	0.8500
Improve RF	0.8450	0.9821	0.9500
DRSN-CS	0.7350	0.9750	0.9333
GCN	0.8133	0.9321	0.9000
WRF-OOB	0.8450	0.9786	0.9500

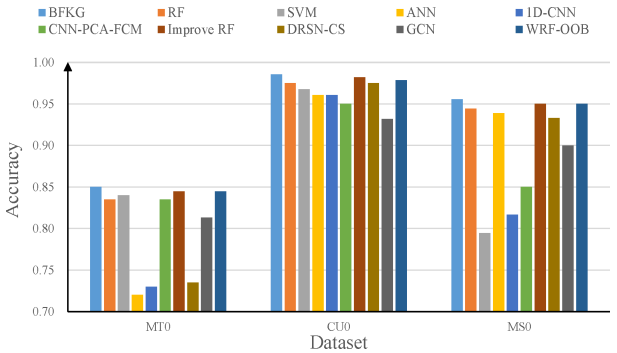


Fig. 6. Experimental results on data MT0, CU0, and MS0. It shows that BFKG is more stable and maintains optimal accuracy compared with other methods.

TABLE IX
COMPARISON OF RESULT OF WHETHER TO ADOPT THE STRATEGY

Dataset	T	Strategy	Test1	Test2	Test3	Test4
MT1	0.75	NO strategy	0.6950	0.7150	0.7400	0.7550
		Data accumulate	0.6950	0.7200	0.7600	0.7700
CU1	0.50	NO strategy	0.9143	0.9286	0.9071	0.9357
		Data accumulate	0.9143	0.9321	0.9393	0.9429
MS1	0.55	NO strategy	0.8833	0.9111	0.8500	0.8667
		Data accumulate	0.8833	0.9167	0.8611	0.8778

C. Case II: Verification of Data Accumulation Effect

In Case II, different thresholds T need to be set for different datasets. The dataset of this case is divided into five parts, one of which is used as the original training set and the other four parts are used as the test set. The experimental group will follow the steps in Section III. First, use the training set to train the model, then add the data worth keeping in this test set to the next training set of the model, record the accuracy of the test set, and repeat the above process until all four test sets are tested. Control group experiment: use the training set to train the model and then test the four test sets to get the accuracy of each test set.

The experimental results of whether the model implements the data accumulation strategy are shown in Table IX, which includes the results on three datasets (CU1, MT1, and MS1). The experimental results show that the classification accuracy of the model after implementing the data accumulation strategy is higher than that of the model without the strategy. Since the training set and other conditions are the same when the model is first trained, the accuracy of Test1 is the same for the results of each dataset, which is also reflected in the table. In Fig. 7, the effectiveness of the data accumulation strategy is more clearly demonstrated using a line chart.

The data accumulation strategy is introduced to other methods. The parameter settings of these methods are the same as in Case I, and the experiment of each method is performed according to the above process. The experimental results of each method implementing the data accumulation strategy are shown in Table X, which includes the results of three datasets (MT1, CU1, and MS1) and various methods.

Experiments show that the classification accuracy of the proposed method after data accumulation is generally better than that of other methods after data accumulation. Among them, the classification accuracy of 1-D CNN and ANN is poor. The reason may be that the data samples used for initial training are too few, and the effect of traditional neural network models depends on the sample size of the training set; otherwise, training overfitting is likely to occur, resulting in poor test results. GCN makes full use of the relationship between samples and thus has a good effect. The proposed method makes full use of the relationship between sample features and fault categories and optimizes the random forest algorithm to obtain the best results.

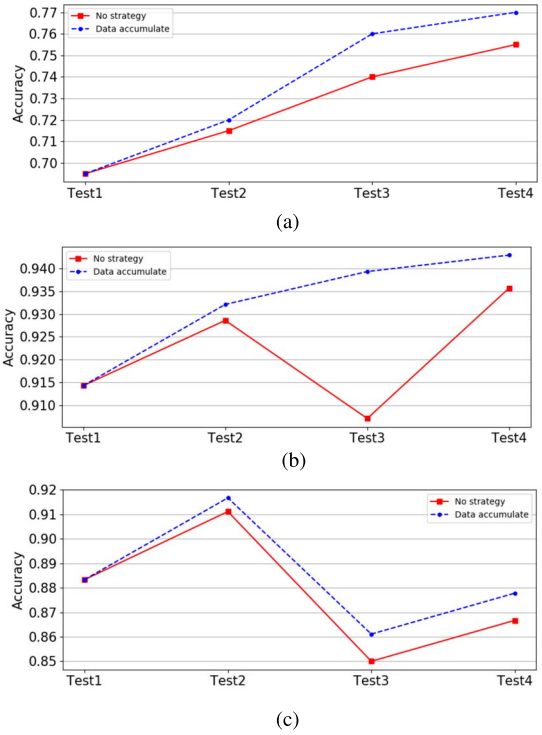


Fig. 7. Experimental results on datasets MT1, CU1, and MS1. (a) Results of case II performed on dataset MT1. (b) Results of case II performed on dataset CU1. (c) Results of case II performed on dataset MS1.

TABLE X
COMPARISON OF OTHER METHODS USING DATA ACCUMULATION STRATEGIES

Dataset	Methods	Test1	Test2	Test3	Test4
MT1	BFKG	0.6950	**0.7200**	**0.7600**	**0.7700**
	SVM	**0.7350**	0.6850	0.6800	0.7500
	ANN	0.4100	0.5400	0.5300	0.6100
	1D-CNN	0.3200	0.3300	0.3250	0.4300
	Improve RF	0.6950	0.6750	0.6800	0.7550
	DRSN-CS	0.6700	0.6450	0.7000	0.7150
	GCN	0.5667	0.4733	0.6000	0.3333
	WRF-OOB	0.6750	0.6850	0.6600	0.7600
CU1	BFKG	0.9143	**0.9321**	**0.9393**	**0.9429**
	SVM	**0.9179**	0.8857	0.8821	0.8924
	ANN	0.7750	0.8036	0.7714	0.7500
	1D-CNN	0.5286	0.6500	0.5964	0.5929
	Improve RF	0.9036	0.9143	0.9214	0.9357
	DRSN-CS	0.9107	0.9071	0.9250	0.9179
	GCN	0.8629	0.8679	0.8750	0.8929
	WRF-OOB	**0.9179**	0.9071	0.9357	0.9357
MS1	BFKG	**0.8833**	**0.9167**	**0.8611**	**0.8778**
	SVM	0.8222	0.8000	0.8222	0.8500
	ANN	0.8056	0.8056	0.8444	0.7722
	1D-CNN	0.6000	0.4722	0.4722	0.5556
	Improve RF	0.8278	0.8667	0.8344	0.8611
	DRSN-CS	0.8389	0.8556	0.8222	0.8167
	GCN	0.7278	0.8056	0.7444	0.7500
	WRF-OOB	0.8222	0.8778	0.8389	0.8722

D. Case III: Robustness of the BFKG

To verify the robustness of the proposed method, an imbalanced dataset experiment was added. Based on the

TABLE XI
DESCRIPTION OF IMBALANCED CWRU DATASETS

Health condition	Label	The ratio of training samples (%)	The ratio of testing samples (%)
N	1-4	80	20
IR	5-12	50	20
OR	13-20	30	20
B	21-28	20	20

TABLE XII
RESULTS OF THE EXPERIMENTS ON THE IMBALANCED DATASET

Dataset	Imbalance
BFKG	**0.9643**
RF	0.9571
SVM	0.9071
ANN	0.9357
1D-CNN	0.6750
CNN-PCA-FCM	0.8643
Improve RF	0.9571
DRSN-CS	0.9321
GCN	0.9107
WRF-OOB	0.9537

CWRU dataset, an imbalanced dataset is constructed by adjusting the number of samples in different states. Instead of having 50 samples per state, the details of the imbalanced dataset are shown in Table XI. The number of training samples for each of the four health states is 40. The number of training samples for each of the eight inner ring fault states is 25. The number of training samples for each of the eight outer ring fault states is 15. The number of training samples for each of the eight ball fault states is 10.

Ten different methods were used to train on this dataset, and the accuracy of each method on the training set was recorded, as shown in Table XII. As can be seen from Table XII, the proposed method still has the highest accuracy with less drop in accuracy compared to the balanced dataset. It shows that the method can effectively deal with the imbalance of training data samples and has strong robustness.

V. CONCLUSION

In this article, a bearing diagnosis method based on KGs is proposed. This method provides a method for constructing KGs, using signal characteristics and fault categories as the nodes of the KGs, and taking the feature-fault correlation as the side of KGs. The KGs constructed by this method can reflect the importance of each feature under different fault categories and lay the foundation for subsequent fault reasoning and diagnosis. An improved random forest algorithm is proposed, which makes full use of feature-fault correlation and has certain advantages over traditional classification algorithms. In addition, a strategy for knowledge accumulation and update of KGs is proposed to make full use of data to realize the gradual expansion of training datasets and improve the accuracy of subsequent data classification.

Two classic datasets and a laboratory dataset were used to experiment with the proposed method. The experimental results verify the superiority of the WRF algorithm and show that the use of this algorithm can make full use of the information provided by the KG. Therefore, it is very suitable and effective to use it as the reasoning algorithm of the KG. The experimental results show that the classification accuracy rate after data accumulation is significantly higher than the classification accuracy rate in the case of no data accumulation, indicating the effectiveness of the data accumulation strategy, which can effectively increase the amount of data in the KGs, thereby improving the inference accuracy. At the same time, this strategy can also provide research ideas for data accumulation of other types of KGs. However, the datasets used in this article are all bearing vibration data of constant speed, and in actual working conditions, the vibration signals of mechanical systems are mostly time-varying signals. Therefore, in the follow-up work, we will try to use the time-varying signal dataset and adopt the feature extraction method for time-varying signals. In addition, the threshold T used in the data accumulation strategy part of this article is roughly drawn up according to the classification effect of the model on the dataset, and it needs to be analyzed in detail and accurately assigned in the follow-up work.

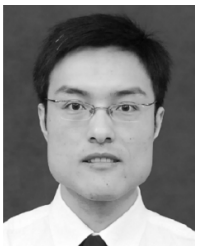
REFERENCES

[1] Z. Yang, Z. Yu, C. Xie, and Y. Huang, "Application of Hilbert–Huang transform to acoustic emission signal for burn feature extraction in surface grinding process," *Measurement*, vol. 47, pp. 14–21, Jan. 2014.
[2] Q. Z. Hu, S. L. Zhang, and S. Yang, "Variable condition bearing fault diagnosis based on time-domain and artificial intelligence," in *Applied Mechanics and Materials*, vol. 203. Wollerau, Switzerland: Trans Tech Publications, 2012, pp. 329–333.
[3] Z. Huo, Y. Zhang, L. Shu, and M. Gallimore, "A new bearing fault diagnosis method based on fine-to-coarse multiscale permutation entropy, Laplacian score and SVM," *IEEE Access*, vol. 7, pp. 17050–17066, 2019.
[4] M. Kang, J. Kim, and J. M. Kim, "An FPGA-based multicore system for real-time bearing fault diagnosis using ultrasampling rate AE signals," *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2319–2329, Apr. 2015.
[5] M. Cui, Y. Wang, X. Lin, and M. Zhong, "Fault diagnosis of rolling bearings based on an improved stack autoencoder and support vector machine," *IEEE Sensors J.*, vol. 21, no. 4, pp. 4927–4937, Feb. 2021.
[6] J. Xiong, Q. Zhang, G. Sun, X. Zhu, M. Liu, and Z. Li, "An information fusion fault diagnosis method based on dimensionless indicators with static discounting factor and KNN," *IEEE Sensors J.*, vol. 16, no. 7, pp. 2060–2069, Apr. 2016.
[7] Y. Benmahamed, M. Teguar, and A. Boubakeur, "Application of SVM and KNN to duval pentagon 1 for transformer oil diagnosis," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 6, pp. 3443–3451, Dec. 2017.
[8] S. E. Pandarakone, S. Gunasekaran, Y. Mizuno, and H. Nakamura, "Application of Naive Bayes classifier theorem in detecting induction motor bearing failure," in *Proc. XIII Int. Conf. Electr. Mach. (ICEM)*, Sep. 2018, pp. 1761–1767.
[9] J. Shi, X. Wu, J. Zhou, and S. Wang, "BP neural network based bearing fault diagnosis with differential evolution & EEMD denoise," in *Proc. 9th Int. Conf. Modelling, Identificat. Control (ICMIC)*, Jul. 2017, pp. 1038–1043.
[10] L. Jiang, Q. Li, J. Cui, and J. Xi, "Rolling bearing fault diagnosis based on higher-order cumulants and BP neural network," in *Proc. 27th Chin. Control Decis. Conf. (CCDC)*, May 2015, pp. 2664–2667.
[11] J. Shen, S. Li, F. Jia, H. Zuo, and J. Ma, "A deep multi-label learning framework for the intelligent fault diagnosis of machines," *IEEE Access*, vol. 8, pp. 113557–113566, 2020.
[12] D. T. Hoang and H. J. Kang, "A motor current signal-based bearing fault diagnosis using deep learning and information fusion," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3325–3333, Jun. 2020.
[13] S. Gao, Z. Pei, Y. Zhang, and T. Li, "Bearing fault diagnosis based on adaptive convolutional neural network with Nesterov momentum," *IEEE Sensors J.*, vol. 21, no. 7, pp. 9268–9276, Apr. 2021.

[14] X. Wang, C. Ma, P. Liu, B. Pan, and Z. Kang, "A potential solution for intelligent energy management–knowledge graph," in *Proc. IEEE Int. Conf. Energy Internet (ICEI)*, May 2018, pp. 281–286.
[15] M. Alshahrani, M. A. Khan, O. Maddouri, A. R. Kinjo, N. Queralt-Rosinach, and R. Hoehndorf, "Neuro-symbolic representation learning on biological knowledge graphs," *Bioinformatics*, vol. 33, no. 17, pp. 2723–2730, Sep. 2017.
[16] X. Yao and B. Van Durme, "Information extraction over structured data: Question answering with freebase," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, 2014, pp. 956–966.
[17] A. Fader, L. Zettlemoyer, and O. Etzioni, "Open question answering over curated and extracted knowledge bases," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2014, pp. 1156–1165.
[18] M. Sheng, Q. Hu, Y. Zhang, C. Xing, and T. Zhang, "A data-intensive CDSS platform based on knowledge graph," in *Proc. Int. Conf. Health Inf. Sci.* Cham, Switzerland: Springer, 2018, pp. 146–155.
[19] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.
[20] C. Bizer *et al.*, "DBpedia—A crystallization point for the web of data," *J. Web Semantics*, vol. 7, no. 3, pp. 154–165, Sep. 2009.
[21] D. Vrandečić and M. Krötzsch, "Wikidata: A free collaborative knowledgebase," *Commun. ACM*, vol. 57, no. 10, pp. 78–85, 2014.
[22] R. Miao, X. Zhang, H. Yan, and C. Chen, "A dynamic financial knowledge graph based on reinforcement learning and transfer learning," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 5370–5378.
[23] Z. Lin, D. Yang, and X. Yin, "Patient similarity via joint embeddings of medical knowledge graph and medical entity descriptions," *IEEE Access*, vol. 8, pp. 156663–156676, 2020.
[24] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
[25] S. Y. Lu, K.-H. Hsu, and L.-J. Kuo, "A semantic service match approach based on WordNet and SWRL rules," in *Proc. IEEE 10th Int. Conf. e-Bus. Eng.*, Sep. 2013, pp. 419–422.
[26] T.-W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Process. Lett.*, vol. 6, no. 4, pp. 87–90, Apr. 1999.
[27] R. Socher, D. Chen, C. D. Manning, and A. Ng, "Reasoning with neural tensor networks for knowledge base completion," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 926–934.
[28] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[29] M. Luo, C. Li, X. Zhang, R. Li, and X. An, "Compound feature selection and parameter optimization of ELM for fault diagnosis of rolling element bearings," *ISA Trans.*, vol. 65, pp. 556–566, Nov. 2016.
[30] Y. Lei, Z. He, Y. Zi, and Q. Hu, "Fault diagnosis of rotating machinery based on multiple ANFIS combination with gas," *Mech. Syst. Signal Process.*, vol. 21, no. 5, pp. 2280–2294, Jul. 2007.
[31] MFPT. *Home-Society for Machinery Failure Prevention Technology*. Accessed: Jul. 13, 2021. [Online]. Available: https://www.mfpt.org/
[32] CWRU. *Bearing Data Center*. Accessed: Jul. 13, 2021. [Online]. Available: https://csegroups.case.edu/bearingdatacenter/home
[33] D. Zhang, Y. Chen, F. Guo, H. R. Karimi, H. Dong, and Q. Xuan, "A new interpretable learning method for fault diagnosis of rolling bearings," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.
[34] X. Dong, G. Li, Y. Jia, and K. Xu, "Multiscale feature extraction from the perspective of graph for hob fault diagnosis using spectral graph wavelet transform combined with improved random forest," *Measurement*, vol. 176, May 2021, Art. no. 109178. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224121001986
[35] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4681–4690, Jul. 2020.
[36] Y. Gao, M. Chen, and D. Yu, "Semi-supervised graph convolutional network and its application in intelligent fault diagnosis of rotating machinery," *Measurement*, vol. 186, Dec. 2021, Art. no. 110084. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S026322412101006X
[37] M. Shahhosseini and G. Hu, "Improved weighted random forest for classification problems," in *Proc. Int. Online Conf. Intell. Decis. Sci.* Cham, Switzerland: Springer, 2020, pp. 42–56.

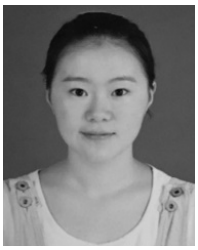
Xiangqu Xiao received the B.S. degree in water conservancy and hydropower engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2019, where he is currently pursuing the Ph.D. degree with the School of Civil and Hydraulic Engineering.

His research interests include nonstationary signal processing and fault diagnosis of rotating machinery.



Chaoshun Li (Member, IEEE) received the B.S. degree in thermal energy and power engineering from Wuhan University, Wuhan, China, in 2005, and the Ph.D. degree in water conservancy and hydropower engineering from the Huazhong University of Science and Technology (HUST), Wuhan, in 2010.

He is currently a Full Professor with the School of Hydropower and Information Engineering, HUST. His research interests include machine learning, intelligent optimization, control theories and applications, and fault diagnosis of rotating machinery.



Jie Huang received the B.S. degree in energy and power engineering from the Kunming University of Science and Technology, Kunming, China, in 2019. She is currently pursuing the Ph.D. degree with the School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan, China.

Her research interests include nonstationary signal processing and fault diagnosis of rotating machinery.

Tian Yu received the B.S. degree from the North China University of Water Resources and Electric Power, Zhengzhou, China, in 2018. She is currently pursuing the Ph.D. degree with the School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan, China.

Her current research interests include mechanical fault diagnosis and reinforcement learning.