A Novel Bearing Fault Classification Method Based on XGBoost: The Fusion of Deep Learning-Based Features and Empirical Features

Jingsong Xie, Zhaoyang Li, Zitong Zhou, and Scarlett Liu, *Member, IEEE*

***Abstract*— The key to intelligent fault diagnosis is to find relevant characteristics with the capability of representing different types of faults. However, the engineering problem is that a few simple empirical features (EFs) cannot obtain high classification accuracy, and complex feature engineering requires strong professional knowledge, which leads to limited applications on a general scale. In addition, intelligent feature extraction and classification methods without prior knowledge cannot guarantee that the model learned the general features used for classification, and its robustness and generalization are not strong when the objects with low-quality training data. Therefore, a fusion method of combining EFs and adaptive features extracted by a deep neural network is proposed. In this method, simple EFs that only need a few professional knowledge are adopted to realize general feature extraction and, hence, maintain the robustness of the model. A modified neural network structure (LiftingNet) is proposed to achieve the adaptive extraction of hidden features for specific objects, for realizing the high precision of bearing fault classification. In order to realize the fusing of EFs and adaptive features, XGBoost is utilized as the final classifier instead of common softmax. The feasibility and validity of the proposed method are verified by two data sets collected from motor bearings at stable work conditions. The experimental results show that the classification accuracy generated by the proposed method is improved. It also can maintain the robust performance on data sets even with various noises.**

***Index Terms*— Bearing fault diagnosis, deep convolutional neural network (DCNN), feature fusion, LiftingNet, XGBoost.**

I. Introduction

BEARING is one of the most common and basic components in industrial mechanical applications. The working condition of the bearing affects the operation status of the equipment [1]. Therefore, fault diagnosis of bearings is critical to ensure the safe and reliable operation of equipment.

Manuscript received July 7, 2020; revised November 15, 2020; accepted November 22, 2020. Date of publication December 7, 2020; date of current version December 30, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61806222 and in part by the Natural Science Foundation of Hunan Province, China, under Grant 2020JJ5771. The Associate Editor coordinating the review process was Anoop Chandrika Sreekantan. *(Corresponding author: Scarlett Liu.)*

Jingsong Xie and Zhaoyang Li are with the School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China.

Zitong Zhou is with the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China.

Scarlett Liu is with the School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China, and also with the School of Mechanical and Manufacturing Engineering, UNSW Sydney, Sydney, NSW 2052, Australia (e-mail: scarlett.liu.academic@gmail.com).

Digital Object Identifier 10.1109/TIM.2020.3042315

Bearing fault diagnosis has been investigated by various works [2]–[6]. In engineering applications, the timely detection of faults through complex signal processing and analysis methods requires comprehensive expertise, which is out of reach for most nonprofessionals who works in the front line in factories. At the same time, practitioners must be able to record and popularize many prior knowledge [7], [8]. Researchers have built many neural network models (such as ANN and SVM) for information processing by using prior fault-sensitive indicators as training data. Those models are then applied for fault classification [9]–[13]. The advantage of them is that it lowered the requirements for practitioners' signal processing capabilities. However, the results are highly correlated with the selected indicators. In addition, it has higher requirements for practitioners to possess prior knowledge about the effective indicators. In recent years, with the development of artificial intelligence, deep learning as a network that relies on multilayer neural networks to mine hidden high-dimensional features for training has been widely used in various industrial applications [14]–[16]. In the field of deep learning, data analysis in the process of fault diagnosis often weakens the expression of the information processing process and emphasizes the relationship between the input signals and fault classes, which can alleviate the nonprofessionals' requirements on data processing and help practitioners find faults in a timely and effective manner [17]–[19].

As to researches on fault diagnosis, scholars either focus on traditional signal decomposition for a better fault frequency band, such as EMD, EWT, and VMD [6], [20]–[23], or believe in the power of AI for fault signal classification, such as AE, DBN, and CNN [16], [24]–[28]. In reality, fault features that can effectively distinguish health conditions are difficult to be composed for complex equipment. At the same time, features extracted by artificial intelligence in each individual training process are mostly related to the form of the network. Also, it is not possible to fully mine all effective fault features. Some researchers studied the ability of feature extraction from deep learning. Lei *et al.* [16] studied the application of AE-based deep learning models in the fault diagnosis of motor bearings. In this study, neural networks were used to autonomously learn the characteristics of the signals, and their research showed that the sparse filtering process is similar to the Gabor filter [16]. Pan *et al.* [29] used multiscale analysis as feature extraction for fault classification.

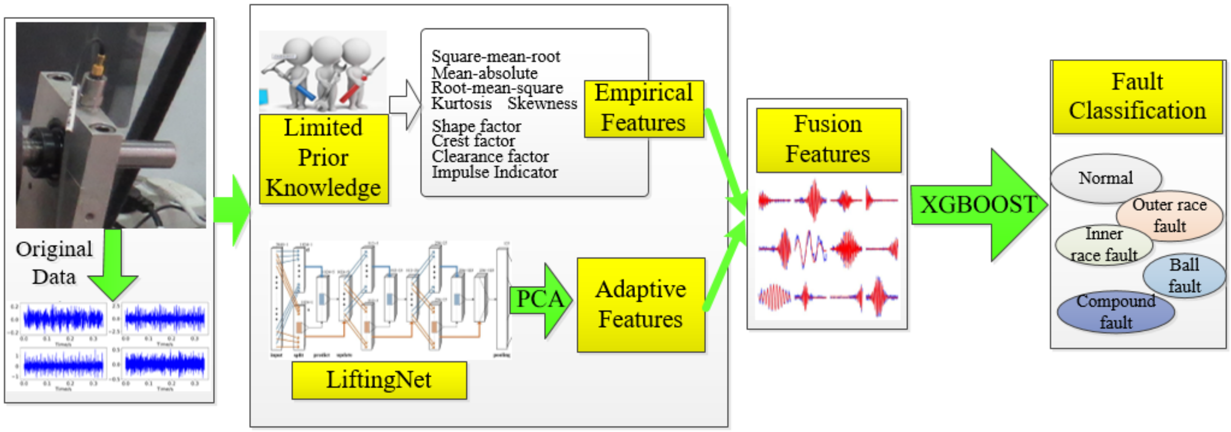

Fig. 1. Proposed model structure.

In recent years, CNN, as a powerful deep learning tool, has been widely used in various classification models [30]–[33]. Many scholars have introduced it into fault diagnosis and produced decent application results. However, CNN is sensitive to noise and interference [34]. Thus, Pan *et al.* [35] proposed a Liftingnet based on CNN classifier, which improves the network on the basis of the second-generation wavelet technology. The lifting framework is also introduced into the built deep neural network model in this article, but, unlike other deep neural network models for fault diagnosis, the XGBoost (eXtreme Gradient Boosting) is used as the final classifier instead of common softmax, as an ensemble learning method [36]. It combines multiple learning models to obtain better results so that the combined model has a stronger generalization ability. XGBoost has been proven to generate high accuracy and fast processing time [37].

Considering practical engineering situations, managers often may not be familiar with professional fault diagnosis knowledge, such as signal decomposition, spectrum analysis, and feature extraction. However, they need to make accurate judgments on the fault status of the equipment. The intelligent feature extraction method can extract hidden features from the original data, but it relies heavily on training data, and the classification result is not stable under different operation conditions. Therefore, this article adopts a feature fusion method that combines a few simple predetermine empirical features (EFs) and adaptive features extracted by deep learning, avoiding the need for strong professional knowledge for feature engineering and the disadvantages of deep learning methods. At the same time, a powerful XGBoost classifier is introduced in the subsequent classification operation for fault classification.

The remainder of this article is organized as follows. In Section II, details of the proposed method are explained. In Section III, three cases of bearing faults and different noise levels are tested to verify the proposed method. Finally, a conclusion is drawn in Section IV.

II. Proposed Approach

A. Model Structure

In this article, a bearing fault diagnosis method, which is based on the fusion of a few EFs relying on limited prior knowledge and the adaptive features extracted from the deep convolutional neural network (DCNN), is established. The designed data processing workflow is demonstrated in Fig. 1. Based on the limited prior knowledge, a few commonly utilized and verified statistical indexes are adopted to extract the general features for bearing fault classification. Adaptive features can be obtained by retraining the modified DCNN (LiftingNet). Then, the principal component analysis (PCA) [38] is utilized to reduce the dimension of generated adaptive features. The EFs are processed by XGBoost as fusing inputs to train the faults classifier.

The structure demonstrated in Fig. 1 contains two modules: the feature extraction module and the classifying module based on the XGBoost model. The feature extraction module can also be divided into two parts. One part is composed of EF. They are classic statistical features extracted from bearing vibration signals using prior knowledge, ensuring to obtain the general faults features. Another part is composed of LiftingNet features (LFs), which is adopted for realizing the extraction of unique hidden features for different objects. Adaptive features are extracted from bearing original vibration signals by retraining the LiftingNet network.

The adaptive network features (LFs) contain redundant information with high dimensions. Thus, it is difficult to get good classification performance by simply putting LF and EF together and sending them into the classification module. Therefore, it is necessary to perform dimension reduction on LFs. PCA is a dimensionality reduction technique that can complete feature fusion while reducing data dimensions; hence, it is utilized to obtain the dimension reduction for final adaptive features.

Those features obtained from the feature extraction module are then fused as the inputs of the XGBoost model. This model is utilized to realize the fusion of EF and LF and, then, finally complete the classification of bearing faults.

B. Feature Extraction

1) Data Preprocessing: Different speeds and lengths of the sampled data affect the final classification. In order to eliminate the influence of these factors on the final test results, a standard data preprocessing produce is presented from two aspects: first, the preprocessed data should keep the

TABLE I
EXPRESSIONS OF NINE ADOPTED EFs

Name	Equation	Name	Equation
Square-mean-root	$p_1 = \left(\frac{1}{N}\sum\sqrt{\lvert x\rvert}\right)^2$	Crest factor	$p_6 = \frac{max(\lvert x\rvert)}{p_3}$
Mean-absolute	$p_2 = \frac{1}{N}\sum\lvert x\rvert$	Shape factor	$p_7 = \frac{p_3}{p_2}$
Root-mean-square	$p_3 = \sqrt{\frac{1}{N}\sum x^2}$	Clearance factor	$p_8 = \frac{max(\lvert x\rvert)}{p_1}$
Kurtosis	$p_4 = \frac{\frac{1}{N}\sum x^4}{p_3^4}$	Impulse Indicator	$p_9 = \frac{max(\lvert x\rvert)}{p_2}$
Skewness	$p_5 = \frac{\frac{1}{N}\sum x^3}{p_3^3}$		

same distribution characteristics as the one from original data; the second is to retain a complete bearing vibration cycle for the sampled data. Therefore, in the data preprocessing step, the original data are sampled into the same vibration cycle length of a bearing, and then, the sampled data are normalized.

The size of the original data is 256 000 points, and the data collection time of the original data is 20 s. The vibration frequency of the bearing is 20 Hz when it reaches a stable state. Therefore, the length of data corresponding to a bearing vibration cycle should be 640 points. The data within the stable state cycle are cut into subsampled data clips, which has 640 points for network training.

The normalization of the data is vital to train a neural network because the normalized data can eliminate the influence of the minimax data on the network training while maintaining the original data distribution. In this article, the following formula is used to normalize the data after subsampling:

$$x_{\text{normlize}}(n) = \frac{x(n) - x_{\text{avg}}}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x(n) - x_{\text{avg}}\right)^2}} \tag{1}$$

$$x_{\text{avg}} = \frac{1}{N}\sum_{n=1}^{N} x(n). \tag{2}$$

2) Empirical Features: The expressions of nine adopted EFs are shown in Table I. The reason we choose them is that they are commonly recognized statistical indicators and they have been verified that they can effectively extract the general features from faults compared with comprehensive indicators used in articles [35]. The extraction and utilization of general features ensure the robustness of the fault classification model.

3) LiftingNet Features: In the case of limited prior knowledge, deep learning methods can be used to extract hidden features. These features do not depend on prior knowledge, but they can provide supplementary fault information for specific objects for improving the accuracy of fault classification. LiftingNet is an efficient DCNN for bearing fault diagnosis, which can extract effective feature information from bearing vibration signals. Its structure is shown in Fig. 2. Therefore, the modified LiftingNet structure, which discards the last fully connected layer of LiftingNet and flattens its final pooling layer, is proposed as an adaptive feature extractor.

4) Features Fusion: The purpose of feature fusion is to combine feature information from different sources and eliminate redundant information in the original features. First, PCA is adopted to reduce the dimension and eliminate redundant information for the high-dimensions outputs of modified LiftingNet, for constructing adaptive features. The feature number is the same as that of the EFs. Then, XGBoost is applied to fuse the EFs and adaptive features. These two complementary features are utilized as the input of the XGBoost model for training a model. Thus, it can achieve the full fusion of general features and special hidden features. In the following cases, the adaptive features extracted by LiftingNet are 125-D, and its dimension was reduced to nine after PCA. The manually extracted features are nine commonly used EFs. Finally, as the input of XGBoost, the dimension of features for fusion is 18.

C. Forward Propagation and Backpropagation of the Modified LiftingNet

In this part, the modified LiftingNet, a powerful tool that was improved from CNN, was applied to extract adaptive features of the vibration signal. The specific structure of LiftingNet can be found in [35]. The cleverness of LiftingNet is that it can learn the data quickly by drawing on the idea of the second-generation wavelet decomposition. With the assistance of the orthogonality of the second-generation wavelet in signal decomposition, the LiftingNet network can better extract the characteristic information in the bearing fault vibration signal. The modified training process of LiftingNet is explained as follows.

First, compared with the linear convolution operation in CNN, LiftingNet uses the prediction and updating steps in the second-generation wavelet to implement the circular

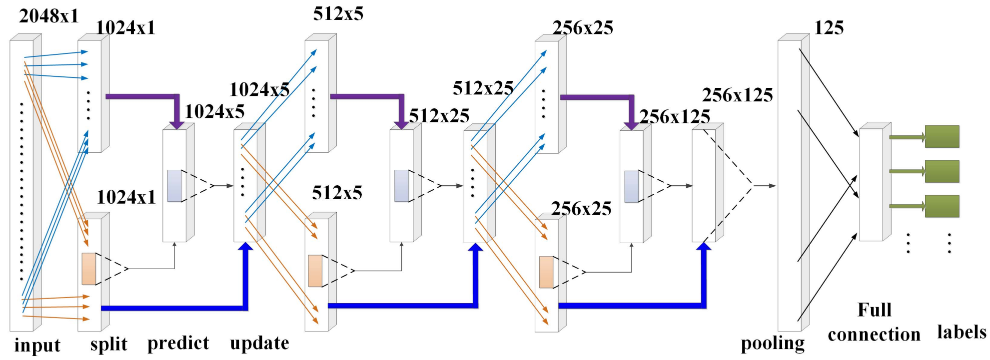

Fig. 2. LiftingNet network structure diagram.

convolution operation, which can improve the learning ability of the network. The specific training process is shown as follows.

In the process of circular convolution, each layer is first decomposed into odd and even sequences

$$d_l^{(k)}(n) = x_l^{(k)}(2n+1) \tag{3}$$
$$s_l^{(k)}(n) = x_l^{(k)}(2n). \tag{4}$$

Then, the prediction and updating algorithms in the SGWT are used for odd and even sequences, respectively. This step is similar to convolution operation. In order to increase the learning ability and deep expression ability of the network, the same steps can be repeated. The abovementioned process can be called circular convolution. The prediction layer is given by (5), and the updating layer is given by (6)

$$x_{l,m}^{(k+2)} = f\left(d_l^{(k)} - g\left(p_{l,m}^{(k+1)}, s_{l,m}^{(k)}\right) + b_{l,m}^{(k+1)}\right) \tag{5}$$
$$x_{l,m}^{(k+3)} = f\left(s_l^{(k)} + g\left(u_{l,m}^{(k+2)}, x_{l,m}^{(k+2)}\right) + b_{l,m}^{(k+2)}\right). \tag{6}$$

Second, after the cyclic convolution operation, the maximum pooling layer and fully connected layer retain the traditional structure of the CNN. Among them, the process of the largest pooling layer is represented by (7), and the fully connected layer is shown by (8)

$$p^{l(i,j)} = \max\left(x^{l(i,j)}\right) \tag{7}$$
$$o = f_2\left(Wx_p + b\right). \tag{8}$$

Third, in the process of backpropagation, a loss function is defined by the following equation:

$$L = \frac{1}{2}\sum(y-o)^2. \tag{9}$$

The goal of the training is to minimize the least-squares error L of the network output, and the error backpropagation algorithm is completed by the BPTT algorithm. The specific derivation process can be found in [35]. The training of the entire network uses a stochastic gradient descent algorithm.

D. Training Process of the XGBoost

In this article, traditional Softmax for classification in the fields of fault diagnosis was replaced with more powerful XGBoost. The XGBoost model is a learning model based on tree embedding. Its basic idea is to combine multiple tree models with low classification accuracy to build a more accurate model. The model is continuously iteratively promoted, and each iteration generates a new tree to fit the residual of the previous tree. The lifting algorithm uses a simple model to fit the data to get a more general result first and then continuously adds simple models to the model to improve the accuracy of the model. Although, in most cases, the XGBoost model is a decision tree with a shallow number of layers, as the number of trees increases, the complexity of the integrated model gradually increases until the complexity of the data itself is approached, and the training result reaches the optimal level. The biggest feature of XGBoost is that it can automatically use the multithreading of the CPU for parallelism, which can improve the algorithm accuracy. For the traditional gradient boosting decision tree (GBDT), only the first-order derivative information is used. When training the nth tree, the residuals of the first $n-1$ trees need to be used, which is difficult to achieve distributed. XGBoost did the second-order Taylor expansion of the loss function and added a regular term overall to the loss function to find the optimal solution to weigh the decline of the loss function and the complexity of the model to avoid outfitting. The main process of XGBoost is shown as follows [36].

First, assume that a model has K decision trees, and the classification result of this model can be shown as follows:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}. \tag{10}$$

The loss function can be given as

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{11}$$

where

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2 \tag{12}$$

T represents the number of leaves, and w is the weight of these leaves;

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i). \tag{13}$$

Then, the loss function can be represented by

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k). \tag{14}$$

Second, the Second-order approximation can be used to quickly optimize the objective in the general setting

$$\mathcal{L}^{(t)} \simeq \sum_{i=1}^{n}\left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(f_t) \tag{15}$$

where $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$ and $h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$. According to (12), (15) can be rewritten as (16) by removing the constant terms $l(y_i, \hat{y}_i^{(t-1)})$

$$\begin{aligned}\tilde{\mathcal{L}}^{(t)} &= \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \gamma T + \frac{1}{2}\lambda\|w\|^2 \\ &= \sum_{j=1}^{T}\left[\left(\sum_{i\in I_j} g_i\right)w_j + \frac{1}{2}\left(\sum_{i\in I_j} h_i + \lambda\right)w_j^2\right] + \gamma T. \end{aligned} \tag{16}$$

Third, differentiate the above formula and make the result equal to 0

$$w_j^* = -\frac{\sum_{i\in I_j} g_i}{\sum_{i\in I_j} h_i + \lambda'}. \tag{17}$$

Then, substituting the optimal weight w_j^* of w_j into the objective function

$$\tilde{\mathcal{L}}^{(t)}(q) = -\frac{1}{2}\sum_{j=1}^{T}\frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_i + \lambda} + \gamma T. \tag{18}$$

It is not possible to enumerate all the tree structures. Therefore, XGBoost uses a greedy algorithm; each time adding a split to the existing leaf node, the information gain is expressed as follows:

$$\mathcal{L}_{\text{split}} = \frac{1}{2}\left[\frac{\sum_{i\in I_L} g_i)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{\sum_{i\in I_R} g_i)^2}{\sum_{i\in I_R} h_i + \lambda} - \frac{\sum_{i\in I} g_i)^2}{\sum_{i\in I} h_i + \lambda}\right] - \gamma. \tag{19}$$

In order to limit the growth of the tree, a threshold is added. When the abovementioned gain is greater than the threshold, the nodes are allowed to split.

III. Case Studies

A. Model for Comparison

1) LiftingNet: LiftingNet is a novel deep learning network that was capable of fault classification; since its introduction last year, it has received a lot of attention. Its uniqueness lies in the combination of the second-generation wavelet decomposition and the convolution operation and pooling in the convolutional neural network. The second-generation wavelet transform (SGWT) has been widely used in fault diagnosis, and its excellent performance is fully confirmed. As one of the deep neural networks widely used in image and semantic classification in recent years, CNN has powerful learning capabilities for the application of deep learning in industrial engineering. Therefore, on the basis of combining the advantages of second-generation wavelet and CNN, LiftingNet was proposed and proved its outstanding performance by two motor bearing data sets.

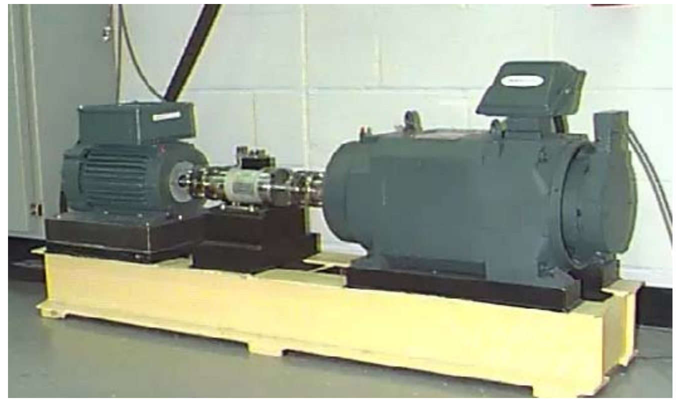

Fig. 3. Experimental apparatus used in the CWRU data set.

2) SVM: Features in the time domain can reflect much information about the working condition of the rolling element bearing. In this article, nine commonly used feature indexes shown in Table I are employed as a feature set for fault diagnosis using SVM. SVM is a typical supervised learning model. In the optimization process, SVM takes into account both empirical risk and structural risk minimization. It has stability and good generalization ability. It is currently widely used to solve engineering problems.

3) XGBoost With Empirical Features: XGBoost itself is a powerful classifier that has received widespread attention in numerous AI competitions since its inception. Therefore, another important comparison method is to classify the vibration data of different fault bearings by using the prior knowledge index as the feature and compare it with the XGBoost method proposed in this article, which combines the prior knowledge and LF as the feature.

B. Case One: CRWU Data

1) Introduction of the Data Set: Bearing fault data produced by Western Reserve University (CRWU data) are widely used in the inspection of various fault diagnosis algorithms, and the experimental apparatus is shown in Fig. 3. In this part, 12k drive end bearing fault data were used to verify the proposed method and other methods. The bearing type is 6205 with a 0.007-in fault diameter, and the motor speed is 1797 rpm.

The CWRU data set contains bearing vibration data for normal bearings and five bearing faults. The five faults are the ball, the inner circle, and the outer circle at 3, 6, and 12 o'clock. Fig. 4 shows the input waveform of the different labels. Both the model presented in this article and the model for comparison are tested on the CWRU data set.

In the experiment of the CWRU data set, we did not adopt the data processing steps proposed in Section II-B1 but referred to the data processing method of LiftingNet's comparison experiment. The data resampling the growth degree of the CWRU data set was 1024, and a total of 465 sample data were obtained. We used 80% of them as the training set to train the model and 20% as the test set to verify the effect of the model.

2) Results of Case One: The fault degree and the fault type of the CWRU data sets are obvious and easy to detect fault signals because it is a standard public data set, which is

TABLE II
EXPERIMENTAL RESULTS BASED ON THE CRWU DATA

Method	normal		BALL		IR		OR3		OR6		OR12		total	
	TP	acc	TP	acc	TP	acc	TP	acc	TP	acc	TP	acc	TP	acc
LiftingNet	145	1.00	111	1.00	120	1.00	134	1.00	120	1.00	137	1.00	767	1.00
SVM	145	1.00	111	1.00	120	1.00	134	1.00	120	1.00	137	1.00	767	1.00
XGBoost	143	0.99	111	1.00	119	0.99	133	0.99	120	1.00	137	1.00	763	0.99
Proposed	145	1.00	111	1.00	119	0.99	134	1.00	120	1.00	137	1.00	766	0.99

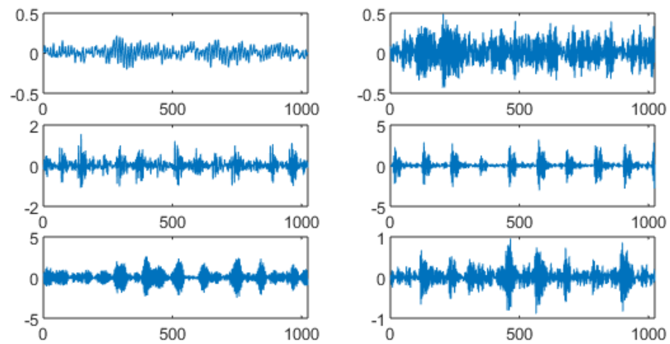

Fig. 4. Original bearing vibration data visualization for each fault type. From top left to bottom right: normal, BALL, IR, OR3, OR6, and OR12.

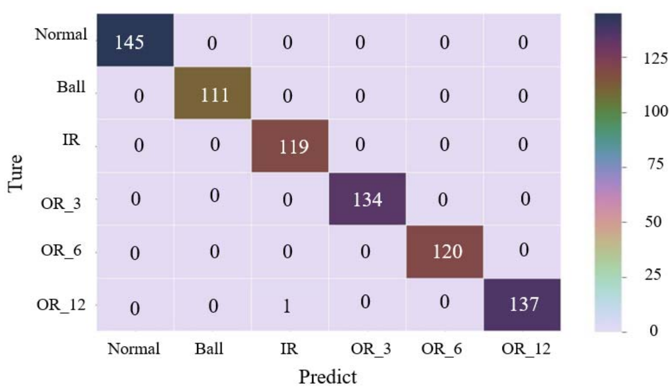

Fig. 5. Confusion matrix of test results on the CWRU data set.

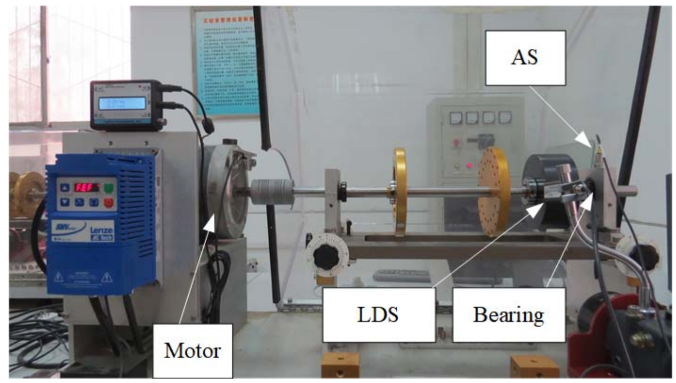

Fig. 6. Snapshot of the SQ experimental apparatus.

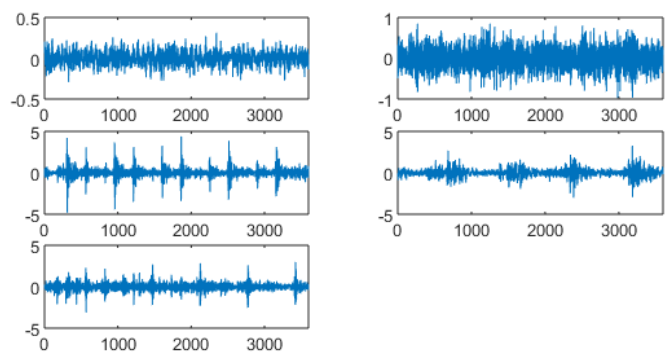

Fig. 7. Original bearing vibration data graph of each failure type. Top left to bottom right: normal, OR, IR, BALL, and joint.

collected from a relatively good experimental condition. The results of each model on the CWRU data set are compared in Table II, and the result confusion matrix of the proposed method on the test set is shown in Fig. 5. In the test of the CWRU data set, compared with LiftingNet and SVM, although the classification accuracy of XGBoost-based classifier is not 100%, the proposed method also achieved an acceptable classification effect (99%). Considering that the fault information in the CRWU vibration data is relatively obvious, SVM can achieve perfect results by using traditional indicators, and LiftingNet can also achieve 100% accuracy by virtue of the advantages of the SGWT. The accuracy of classification is slightly lower than 100% because XGBoost is difficult to adjust parameters during training.

C. Case Two: Spectral Quest Data

1) Introduction of the Data Set: In this section, vibration data are collected from five types of bearing faults. There are normal bearing, outer race fault bearing, inner race fault bearing, ball fault bearing, and compound fault bearing. The experimental platform is a machinery fault simulator shown in Fig. 6, and the testing subject is an ER-16K bearing on the right of the shaft. A laser displacement sensor (LDS) is used to record speed pulses, and three acceleration sensors (ASs) are stuck on the bearing block in the vertical direction. The control of motor speed is executed by a computer. The bearing vibration data of different fault types are shown in Fig. 7.

2) Result of Case Two: The results of each model on the data set collected in spectral quest (SQ) experiments are compared in Table III. In the abovementioned experiments, as the fault types in SQ experiments are not obvious as those in the CWRU data set, the classification accuracies of all the mentioned methods are reduced. However, the classification accuracy of the proposed method is 92%, which is significantly higher than that of the other methods, such as XGBoost with

TABLE III
EXPERIMENTAL RESULTS ON THE DATA SET COLLECTED IN SQ EXPERIMENTS

Method	normal		IR		OR		BALL		JOINT		total	
	TP	acc	TP	acc	TP	acc	TP	acc	TP	acc	TP	acc
LiftingNet	252	0.95	276	0.88	289	0.91	260	0.79	241	0.92	1318	0.89
SVM	252	0.95	268	0.85	289	0.89	274	0.83	223	0.89	1306	0.88
XGBoost	253	0.98	287	0.88	299	0.93	277	0.88	245	0.91	1361	0.91
Proposed	253	0.98	286	0.91	304	0.95	279	0.86	249	0.93	1371	0.92

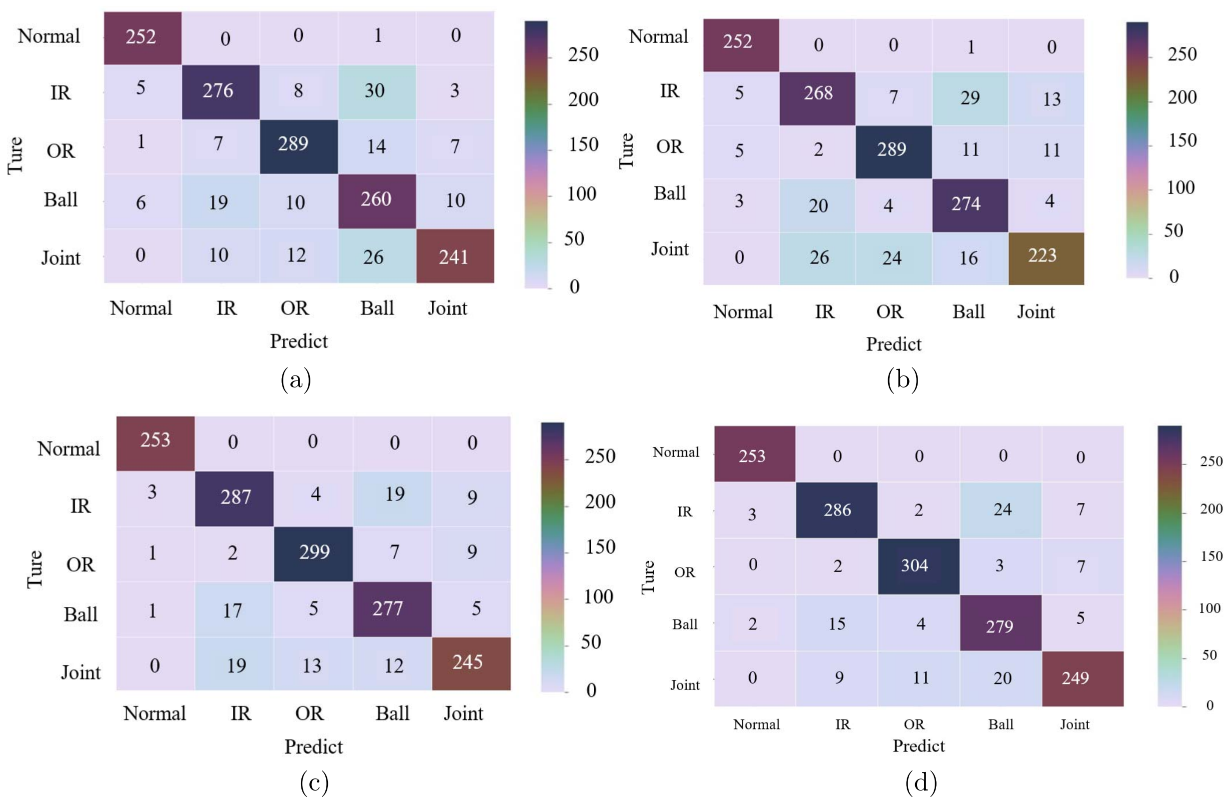

Fig. 8. Comparisons of confusion matrix in case two. (a) LiftingNet. (b) SVM. (c) XGBoost. (d) Proposed method.

EFs: 91%, SVM: 88%, and Liftingnet: 89%. In addition, the confusion matrix of the proposed method and three other methods are compared in Fig. 8. Those matrices illustrate that the proposed method has advantages in classification accuracy and fault recall rate. The comparison results show that the proposed method has a higher classification accuracy in the case where the fault type is not obvious, which verifies the effectiveness of the proposed feature fusion method. In engineering practice, combining limited prior knowledge with deep learning can achieve good results that can also provide better decision-making reference for nonprofessionals and reduce their learning costs. In subsequent discussions, the robustness of the proposed method to noise to illustrate the superiority of the proposed method is further verified.

D. Case Three: Noise Influence on CRWU Data

In this part, we discuss the influence of noise on this method and compare it with other methods. We used the data set of case one in which all the abovementioned methods have good performance, as shown in Section III-B, when no additional noise is added. We added the noise with random normal distribution after the weighted sum of randomly selected samples to generate artificial sample data for classification and evaluation of the impact of the model on noise.

For each artificial sample data, our generation method is to randomly select two real data x_1 and x_2 under the label of the same type, randomly assign them two weights between 0 and 1, w_1 and w_2, and satisfy $w_1 + w_2 = 1$. Use (20) to generate new artificial data with different noises

$$x_{\text{new}} = \left[1 + \gamma N(0, 1)\right](w_1 x_1 + w_2 x_2) \tag{20}$$

where $N(0, 1)$ is a random vector of the normal distribution with variance 0 and mean value 1, and its length is consistent with that of the samples. γ is the scale parameter controlling the noise level, and its value is 0.05, 0.2, 1.0, and 2.0, which means that, through the transformation of (20), it is equivalent to multiplying a random scaling coefficient of normal distribution with a mean value of 1 and a variance

TABLE IV
EXPERIMENTAL RESULTS ON THE CWRU DATA SET BY ADDING DIFFERENT NOISES

method	original data	noise=0.05	noise=0.2	noise=1	noise=2
LiftingNet	1.00	0.9935	0.9126	0.8644	0.7119
SVM	1.00	0.9984	0.9917	0.9278	0.9021
XGBoost	0.9985	0.9961	0.9930	0.9484	0.9403
Proposed	0.9970	0.9980	0.9937	0.9547	0.9507

of γ for each sample point to simulate noises. These samples can simulate the bearing vibration data in engineering application environment with different noise intensity, to verify that the proposed method has better fault identification accuracy for bearings in different application scenarios. In each test, we generated 500 artificial sample data for each fault type. The test results are illustrated in Table IV.

In case one, as the fault types are obvious, all the mentioned methods can achieve a good classification performance. However, in this case, by introducing noise in the data set, the accuracies of all these methods are reduced, but our methods still maintain relatively higher accuracies than SVM and Liftingnet. As shown in Table IV, in the abovementioned noise range, the classification accuracy of the proposed method has the smallest reduction and is higher than 95%, which means that the proposed method is more stable than other mentioned methods. The reason for this may be is that the fusion of the deep neural network-based features and the EFs can obtain both the fixed and verified important strong professional knowledge, and the latent information adapts to the change of data.

IV. CONCLUSION

This article presents a new method based on XGBoost for bearing fault classification, which fuses two types of features to obtain the general features for bearing faults and hidden features for special objects. The features used to train the XGBoost model classifier is the fusion of a few simple EFs and the adaptive features extracted by LiftingNet. The accuracy and robustness of the classification are verified with two motor bearing data sets. The same accuracy in case one (an easy case) verified the feasibility of the proposed method. The high accuracy in case two and high robustness to noise in case three verified the superiority of the proposed method. The research results demonstrate that it can improve the accuracy and antinoise effect of the model to some extent by the fusion of a few EFs and the hidden features extracted by the modified LiftingNet.

In addition, the proposed features fusion method avoids the need for strong professional knowledge for feature engineering. It also overcomes the shortcomings that intelligent methods can guarantee a model that is learned from general features to classify faults accurately under the condition of lacking prior knowledge. No mention that the robustness and generalization of that model are not strong when the objects have low-quality training data.

However, the experimental data for cases 2 and 3 were collected in a stable working environment. The future work will be the focus on applying the proposed method under variable working conditions.

ACKNOWLEDGMENT

The authors would like to thank the CV-MLI Laboratory, School of Traffic & Transportation Engineering, Central South University, Changsha, China, for providing all computational platforms to process data.

REFERENCES

[1] X. Wang, Y. Zi, and Z. He, “Multiwavelet denoising with improved neighboring coefficients for application on rolling bearing fault diagnosis,” *Mech. Syst. Signal Process.*, vol. 25, no. 1, pp. 285–304, 2011.

[2] V. Rai and A. Mohanty, “Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform,” *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2607–2615, 2007.

[3] Y. Lei, J. Lin, Z. He, and Y. Zi, “Application of an improved kurtogram method for fault diagnosis of rolling element bearings,” *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1738–1749, Jul. 2011.

[4] S. Lu, Q. He, and J. Zhao, “Bearing fault diagnosis of a permanent magnet synchronous motor via a fast and online order analysis method in an embedded system,” *Mech. Syst. Signal Process.*, vol. 113, pp. 36–49, Dec. 2018.

[5] Z. Feng, Y. Zhou, M. J. Zuo, F. Chu, and X. Chen, “Atomic decomposition and sparse representation for complex signal analysis in machinery fault diagnosis: A review with examples,” *Measurement*, vol. 103, pp. 106–132, Jun. 2017.

[6] Y. Lei, J. Lin, Z. He, and M. J. Zuo, “A review on empirical mode decomposition in fault diagnosis of rotating machinery,” *Mech. Syst. Signal Process.*, vol. 35, nos. 1–2, pp. 108–126, Feb. 2013.

[7] Z. Gao, C. Cecati, and S. X. Ding, “A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.

[8] Z. Gao, S. X. Ding, and C. Cecati, “Real-time fault diagnosis and fault-tolerant control,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3752–3756, Jun. 2015.

[9] A. Widodo and B.-S. Yang, “Support vector machine in machine condition monitoring and fault diagnosis,” *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, Aug. 2007.

[10] A. Hajnayeb, A. Ghasemloonia, S. E. Khadem, and M. H. Moradi, “Application and comparison of an ANN-based feature selection method and the genetic algorithm in gearbox fault diagnosis,” *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10205–10209, Aug. 2011.

[11] R. Liu, B. Yang, E. Zio, and X. Chen, “Artificial intelligence for fault diagnosis of rotating machinery: A review,” *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.

[12] J. Li, X. Yao, X. Wang, Q. Yu, and Y. Zhang, “Multiscale local features learning based on BP neural network for rolling bearing intelligent fault diagnosis,” *Measurement*, vol. 153, Mar. 2020, Art. no. 107419.

[13] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, and C. Li, “Fault diagnosis in spur gears based on genetic algorithm and random forest,” *Mech. Syst. Signal Process.*, vols. 70–71, pp. 87–103, Mar. 2016.

[14] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[15] M. S. Hossain, M. Al-Hammadi, and G. Muhammad, “Automatic fruit classification using deep learning for industrial applications,” *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 1027–1034, Feb. 2019.

[16] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.

[17] G. Hu, H. Li, Y. Xia, and L. Luo, "A deep Boltzmann machine and multi-grained scanning forest ensemble collaborative method and its application to industrial fault diagnosis," *Comput. Ind.*, vol. 100, pp. 287–296, Sep. 2018.

[18] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 272, pp. 619–628, Jan. 2018.

[19] T. Chen, Z. Wang, X. Yang, and K. Jiang, "A deep capsule neural network with stochastic delta rule for bearing fault diagnosis on raw vibration signals," *Measurement*, vol. 148, Dec. 2019, Art. no. 106857.

[20] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[21] J. Pan, J. Chen, Y. Zi, Y. Li, and Z. He, "Mono-component feature extraction for mechanical fault diagnosis using modified empirical wavelet transform via data-driven adaptive Fourier spectrum segment," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 160–183, May 2016.

[22] Z. Li, J. Chen, Y. Zi, and J. Pan, "Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive," *Mech. Syst. Signal Process.*, vol. 85, pp. 512–529, Feb. 2017.

[23] Y. Wang, R. Markert, J. Xiang, and W. Zheng, "Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system," *Mech. Syst. Signal Process.*, vols. 60–61, pp. 243–251, Aug. 2015.

[24] T. Pan, J. Chen, J. Pan, and Z. Zhou, "A deep learning network via shunt-wound restricted Boltzmann machines using raw data for fault detection," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4852–4862, Jul. 2020.

[25] H. Shao, H. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mech. Syst. Signal Process.*, vol. 102, pp. 278–297, Mar. 2018.

[26] S.-S. Zhong, S. Fu, and L. Lin, "A novel gas turbine fault diagnosis method based on transfer learning with CNN," *Measurement*, vol. 137, pp. 435–453, Apr. 2019.

[27] R. Yan, X. Chen, P. Wang, and D. M. Onchis, "Deep learning for fault diagnosis and prognosis in manufacturing systems," *Comput. Ind.*, vol. 110, pp. 1–2, Sep. 2019.

[28] D. Cabrera *et al.*, "Bayesian approach and time series dimensionality reduction to LSTM-based model-building for fault diagnosis of a reciprocating compressor," *Neurocomputing*, vol. 380, pp. 51–66, Mar. 2020.

[29] T. Pan, J. Chen, Z. Zhou, C. Wang, and S. He, "A novel deep learning network via multiscale inner product with locally connected feature extraction for intelligent fault detection," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5119–5128, Sep. 2019.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[31] A. Kumar, C. P. Gandhi, Y. Zhou, R. Kumar, and J. Xiang, "Improved deep convolution neural network (CNN) for the identification of defects in the centrifugal pump using acoustic images," *Appl. Acoust.*, vol. 167, Oct. 2020, Art. no. 107399.

[32] J. Zhang, S. Yi, G. Liang, G. Hongli, H. Xin, and S. Hongliang, "A new bearing fault diagnosis method based on modified convolutional neural networks," *Chin. J. Aeronaut.*, vol. 33, no. 2, pp. 439–447, 2020.

[33] B. Zhao, X. Zhang, H. Li, and Z. Yang, "Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions," *Knowl.-Based Syst.*, vol. 199, Jul. 2020, Art. no. 105971.

[34] S. Zheng, Y. Song, T. Leung, and I. Goodfellow, "Improving the robustness of deep neural networks via stability training," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4480–4488.

[35] J. Pan, Y. Zi, J. Chen, Z. Zhou, and B. Wang, "LiftingNet: A novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification," *IEEE Trans. Ind. Electron.*, vol. 65, no. 6, pp. 4973–4982, Jun. 2018.

[36] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.

[37] S. R. Mousa, P. R. Bakhit, and S. Ishak, "An extreme gradient boosting method for identifying the factors contributing to crash/near-crash events: A naturalistic driving study," *Can. J. Civil Eng.*, vol. 46, no. 8, pp. 712–721, Aug. 2019.

[38] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.

Jingsong Xie was born in Anren, Hunan, China, in 1989. He received the B.S. degree from the School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an, China, in 2013, and the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, in 2018.

He then joined the School of Traffic and Transportation Engineering, Central South University, Changsha, China, as a Lecturer. His research interests include intelligent diagnosis of a mechanical fault, mechanical signal processing, and vibration analysis.

Zhaoyang Li received the B.S. degree in traffic and transportation engineering from Central South University, Changsha, China, in 2019, where he is currently pursuing the master's degree in traffic and transportation engineering.

His research interests include machine learning, fault diagnosis of mechanical equipment, signal processing, passive safety protection of locomotive, and finite element analysis.

Zitong Zhou received the B.S. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2014, where he is currently pursuing the Ph.D. degree in mechanical engineering.

His research interests include condition monitoring of mechanical equipment and mechanical signal processing.

Scarlett Liu (Member, IEEE) is currently a Senior Lecturer in vision sensing for field robotics discipline with the School of Traffic and Transportation Engineering, Central South University, Changsha, China. She is also a Research Associate with the School of Mechanical and Manufacturing Engineering, UNSW Sydney, Australia. Her research interests are focused on robotics with experience in robotic vision, sensing, and environment reconstruction. She does a theoretical investigation on unsupervised data structure learning from the perspective of vision perception and sensing for autonomous systems, which supports decision-making and autonomous system control.