Spur Gear Fault Diagnosis Using a Multilayer Gated Recurrent Unit Approach With Vibration Signal

YING TAO[1], (Member, IEEE), XIAODAN WANG[1], RENÉ-VINICIO SÁNCHEZ[2],
SHUAI YANG [1], (Member, IEEE), AND YUN BAI [1], (Member, IEEE)
[1]National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400067, China
[2]GIDTEC Universidad Politécnica Salesiana, Cuenca EC010102, Ecuador

Corresponding author: Yun Bai (yunbai@foxmail.com)

ABSTRACT The fault diagnosis of the gearbox is a complex and important work. In this paper, a multilayer gated recurrent unit (MGRU) method is proposed for spur gear fault diagnosis, that is, three-layer gated recurrent unit (GRU). The vibration signals are firstly monitored on the test bench, and then extracted in both time domain and time-frequency domain. Finally, MGRU is used to learn representation and classification. The MGRU can improve the representation of information and identify the features of fault types more precisely with the increasing number of layers. The proposed method was tested by two spur gears with 10 state modes. To evaluate the method's classification accuracy, four methods were utilized for comparison, i.e., the GRU, long short-term memory (LSTM), multilayer LSTM (MLSTM), and support vector machine (SVM), respectively. In addition, the separability and robustness analysis are also discussed for the proposed MGRU performance. All of the results exhibited that the proposed MGRU approach is effective for spur gear fault diagnosis.

INDEX TERMS Fault diagnosis, vibration, gated recurrent unit, separability, robustness.

I. INTRODUCTION

With the improvement of automation in modern production, the demand for equipment is increasing sharply, and the maintenance of equipment is also paid more attention [1]. As an important part of mechanical equipment, gearbox is a crucial connection and transmission device. In the gearbox, the gear failure rate is the highest [2], such as wear, pitting and broken teeth [3]–[5]. These faults may lead to the interruption of the production of machinery and equipment, resulting in huge losses and even human casualties [1]. Therefore, the study of gear fault monitoring and diagnosis is extremely critical.

In the early stage of the development of statistical science, researchers and engineers can only make some assumptions about the background distribution of a small sample, such

The associate editor coordinating the review of this manuscript and approving it for publication was Dong Wang.

as independent normal distribution, and then establish some hypothetical mathematical models. But these methods cannot deal with complex big data. With the development of computational systems, scholars have researched many computational methods (or data-driven model) to deal with a large number of problems. For instance, artificial decision trees [6], random forests [7], support vector machines (SVM) [8] and other algorithmic models [9], [10] have been reported and widely used in this field.

As reported in literature [11], a hierarchical clustering selection based weighted random forests scheme is proposed for fault classification in complex industrial processes. Literature [12] presented a novel signal processing scheme, bandwidth empirical mode decomposition, and adaptive multi-scale morphological analysis for early fault diagnosis of rolling bearings. In these methods, the SVM has been successfully applied for fault diagnosis because of its accurate

classification ability [13]–[15]. The viewpoint of HVCBs based on an adaptive kernel principal component analysis and the SVM method is put forward in literature [16].

Data sequence information collected by sensors is very important for describing machine state. However, the machine learning methods mentioned above may not capture the intrinsic relationship between data and state [17]. As a branch of machine learning model, deep learning has great advantages in this respect [18], [19]. It mainly extracts hierarchical representation from input data by establishing a depth neural network of multi-layer non-linear transformation. By connecting layers with layers, the conversion from input value to output value can be achieved, and the information behind the data can be explored sufficiently. Literature [20] proposed a deep classifier model, which indicated the superiority of deep learning.

In deep learning, long short-term memory (LSTM) model with memory function has received a lot of attention and research [21], [22]. But the LSTM has little potential for more complex tasks. To solve this problem, gated recurrent unit (GRU), proposed by Cho [23] in 2014, can better handle large training data. The GRU synthesizes a single update gate with the forgetting gate and the input gate. The final model is simpler and faster than the standard LSTM model [24]. It has been used in video recognition [25], atmosphere quality management [26], and some other fields. In the machine health monitoring, [27] experimented the local feature-based GRU on machine health monitoring tasks to verify the effectiveness.

However, these data-driven approaches are limited to specific research sample style to some extent. Therefore, the study of different data characteristics of samples can more accurately establish a high-performance analysis model [28]. For fault diagnosis, extracting common time, frequency and time-frequency domain analysis features and using them as input of the model can achieve better data fusion [29]. Follow these principles, the structural steps of this paper are divided into the following three phases: (1) signal acquisition. Acquisition of useful signals is the precondition. Generally, the signals of vibration [30], sound [31], electric current [32] and some other signals are collected in most cases [33], [34]. A fault monitoring test bench is constructed in this paper, and vibration signals are collected, which is also the most commonly used signal acquisition method for researchers [35]. (2) Signal processing. Manual extraction of some representative features. (3) Fault diagnosis using appropriate models. In this paper, the author uses the deep learning method to construct the neural network gear fault diagnosis, and proposes a gear fault diagnosis method based on multilayer gated recurrent unit (MGRU) model.

The main original contributions of this research are presented as follows:

1) The multi-layer neural network model has a very strong learning ability. The feature data processed by the multi-layer model can simulate the original data better, which is conducive to the realization of classification problems. The advantage of this method is that it can express features more deeply and has stronger function simulation ability. Each layer of neurons learns the more abstract representation from the former layer, thus gaining better discrimination and classification ability. At the same time, due to adding a small number of layers and parameters, the computational complexity is within the acceptable range.

2) For gear fault diagnosis, this paper compared the LSTM, the multilayer long short-term memory (MLSTM), the GRU, the MGRU and the SVM models respectively. The experimental analysis of these models shown that the MGRU network can effectively improve the accuracy of gear fault diagnosis.

The rest of this paper is organized as follows. Section II describes the modeling methods used in this paper. Section III illustrates experimentation test bench, data set and feature extraction method. Section IV gives the experimental results of the proposed model and compares the other machine learning algorithm for fault diagnosis. Finally, conclusions are given in Section V.

II. METHODOLOGIES

In this section, the systematic methodologies are described in detail. The first subsection introduces the feature extraction method, the classifier MGRU for the modes' fault diagnosis is described in the second subsection, the third subsection summarizes our experimental steps, and the compared approaches are drawn in the last subsection.

A. FEATURE EXTRACTION

Sensors can only be placed outside the shell, and the signals collected are the synthesis of vibration information. Therefore, it is difficult to distinguish whether a machine has a fault or not only by extracting the time-domain features. Wavelet Packet Transform (WPT) can intuitively and effectively describe the fault characteristics of signals [36]. The process of extracting characteristic parameters of WPT energy spectrum is as below.

Firstly, choose suitable mother wave to decompose the vibration signal with n-layer wavelet packet, and the coefficient vectors are obtained as follows:

$$[x_n^0 x_n^1 x_n^2 \ldots, x_n^{2^n-1}] \tag{1}$$

where $x_n^i (i = 0, 1, 2, \ldots, 2^n - 1)$ denotes the coefficients of the i-th node of the n-th layer decomposed.

Secondly, when the wavelet packet decomposition coefficient is reconstructed, the reconstructed total signal S can be expressed as:

$$S = \sum_{i=0}^{2^n-1} S_n^i \tag{2}$$

where S_n^i denotes signal of the i-th node in the n-th layer.

TABLE 1. Description of 12 time-domain statistical features.

Time-domain features	Formula	Description
Average value	$A_1(x) = \dfrac{1}{N}\sum\limits_{i=1}^{N} x(i)$	$A_1(x)$: measures the average change of signals.
Standard deviation	$A_2(x) = \sqrt{\dfrac{1}{N}\sum\limits_{i=1}^{N}\left\| x(i) - A_1(x)\right\|^2}$	$A_2(x)$: measures the energy evolution of a signal.
Skewness	$A_3(x) = \dfrac{\dfrac{1}{N}\sum\limits_{i=1}^{N}(x(i) - A_1(x))^3}{[\dfrac{1}{N}\sum\limits_{i=1}^{N}x^2(i)]^{\frac{3}{2}}}$	$A_3(x)$: measures the asymmetry of vibration Signals.
Peak-to-peak value	$A_4(x) = \max(x(i)) - \min(x(i))$	$A_4(x)$: measures the asymmetry of the data about its mean value.
Variance	$A_5(x) = A_2(x)^2$	$A_5(x)$: measures the fluctuation degree of signal deviation from center trend.
Kurtosis	$A_6(x) = \dfrac{\sum\limits_{i=1}^{N}(x(i) - A_1(x))^4}{N \times A_3(x)^4}$	$A_6(x)$: measures the peakedness, smoothness, and the heaviness of tail in a data set.
Root mean square	$A_7(x) = \sqrt{\dfrac{1}{N}\sum\limits_{i=1}^{N}x^2(i)}$	$A_7(x)$: measures the energy evolution of a signal.
Waveform factor	$A_8(x) = \dfrac{A_7(x)}{\dfrac{1}{N}\sum\limits_{i=1}^{N}x(i)}$	$A_8(x)$: a dimensionless index which is more sensitive to the operation state of the unit.
Peak factor	$A_9(x) = \dfrac{\max(x(i))}{A_7(x)}$	$A_9(x)$: measures how extreme the peaks are in a waveform.
Kurtosis factor	$A_{10}(x) = \dfrac{A_6(x)}{A_7(x)^4}$	$A_{10}(x)$: measures the shock characteristics in vibration signals.
Impulse factor	$A_{11}(x) = \dfrac{\max(x(i))}{\dfrac{1}{N}\sum\limits_{i=1}^{N}x(i)}$	$A_{11}(x)$: measures the shock in the signals.
Margin factor	$A_{12}(x) = \dfrac{\max(x(i))}{A_1(x)}$	$A_{12}(x)$: measures the wear of machinery and equipment.

Then the energy values corresponding to each sub-band are calculated.

$$E_n^i = \int \left| S_n^i(t)\right|^2 dt = \sum_{k=1}^{N} |x_{ik}|^2 \qquad (3)$$

where x_k^i represents the amplitude of the reconstructed signal S_n^i, and N denotes the number of vibration signal samples.

Finally, the eigenvector of energy spectrum F is finally constructed.

$$\begin{cases} E = \sum\limits_{i=0}^{2^n-1} (E_n^i)^2 \\ F = \left[\dfrac{E_n^0}{\sqrt{E}}, \dfrac{En^1}{\sqrt{E}}, \dots, \dfrac{E_n^{2^n-1}}{\sqrt{E}} \right] \end{cases} \qquad (4)$$

This paper uses common time domain features and the WPT to extract features in different time and time-frequency domains, thus providing an effective method for more complex fault diagnosis. For the signals measured in the accelerometers, the 20 considered features are used. There are 12 time-domain statistical features summarized in TABLE 1 and 8 time-frequency features, extraction of time-frequency features by WPT. In the experiment of the vibration signal, the WPT with "daubechies 1" (mother wave function) is applied to perform 3-levels decomposition, and results in 8 decomposed frequency bands. Through calculation, the energy sum for each frequency band can be obtained, which represents the 8 features of vibration signal in time-frequency domain ($w_1 - w_8$).

B. MGRU

Gated Recursive Unit (GRU) [23] is a model of the classification and prediction algorithm variation from long short term memory network (LSTM) [24]. The core idea of the LSTM and GRU is "cell state". It controls the ability of information to flow to the cell state by a structure called "gate". A gate includes a sigmoid neural network layer and a multiplication operation, which is shown in FIGURE 1.

The values between 0 and 1 are outputs by the "Sigmoid" layer, where 0 stands for "no quantity allowed to pass" and 1 means "any quantity allowed to pass". The LSTM has three gates to protect and control the "unit" state in order to maintain long-term dependence on information. Each black carries an entire vector, and the circle represents a point-wise
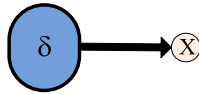
FIGURE 1. The illustrations of a "gate".

FIGURE 2. The illustrations of GRU model.

operation (The function values of each point in the definition field are calculated separately). The main steps of LSTM are as follows:

1) Forget gate f_t. It's used to decide what information to throw away from a cell. For the t-th sample of a set of data, the current input x_t and neuron information l_t are read by this layer, and the discarded information is determined by f_t.

$$f_t = \delta(W^f x_t + U^f l_{t-1}) \tag{5}$$

where the δ is activation function, W^f is the weight matrices of f_t gate, and the U^f is parameter matrix.

2) Input gate i_t. It is used to identify new information stored in cells state, including input gate layer and a new memory cell. These two layers represent the values i_t to be updated and add a new candidate value vector d'_t, respectively. That is, adding new ones to the cell state to replace the old ones that need to be forgotten.

$$i_t = \delta(W^i x_t + U^i l_{t-1}) \tag{6}$$
$$d'_t = \tanh(W^{d'} x_t + U^{d'} l_{t-1}) \tag{7}$$

where the *tanh* is activation function, W^i, $W^{d'}$ are weight matrices of i_t and d'_t gates, respectively. The U^i and $U^{d'}$ are parameter matrixes.

3) Output gate m_t. It is used to update the state of old cells. That's to turn d_{t-i} into d_t.

$$d_t = f_t \circ d'_{t-1} + i_t \circ d'_t \tag{8}$$
$$m_t = \delta(W^m x_t + U^m l_{t-1}) \tag{9}$$

where the $\circ$ denotes the element-wise product, W^m is the weight matrices of m_t gate, and the U^m is parameter matrix.

4) Output the final result l_t. It is based on cell state. Specific steps are as follows:

$$l_t = m_t \circ \tanh(d_t) \tag{10}$$

The GRU is a very popular variant of the LSTM, and the same gate control mechanism is also used in the GRU. It can overcome the problem of gradient disappearance. But there are some differences in the GRU. That is, the forget gate f_t and input gate i_t in the LSTM are synthesized into an update gate. Similarly, cell state and hidden state are mixed. Hence the GRU has only two gates: update gate z_t and reset gate r_t. Compared with the LSTM, the structure is simpler and the calculation of the GRU is much less. Its structure is explained in FIGURE 2.
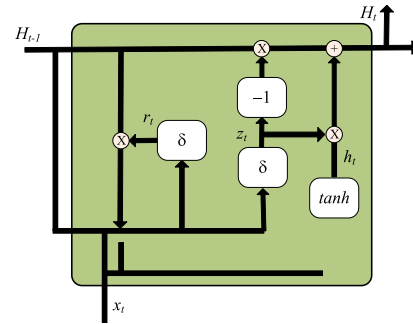
The update gate z_t is used to control the extent of the state information between the previous moment and the current state. The reset gate r_t is used to control the degree of ignoring the state information of the previous moment. For an input vector x_t, conversion functions in the GRU hidden elements are given as follows:

$$z_t = \delta(W^z x_t + V^z H_{t-1}) \tag{11}$$
$$r_t = \delta(W^r x_t + V^r H_{t-1}) \tag{12}$$
$$h_t = \tanh(W^c x_t + V^c(r_t \circ h_{t-1})) \tag{13}$$
$$H_t = (1 - z_t) \circ H_{t-1} + z_t \circ h_t \tag{14}$$

where W^z, W^r are weight matrices of z_t and r_t gates, respectively, W^c is the weight matrix of the output state, H_{t-1} is the input data at time t- 1, $\circ$ denotes the element-wise product, h_t and H_t are candidate states and output states at time t, δ and *tanh* are activation functions for update gate and reset gate, and V^z, V^r are parameter matrices and vector.

For the neural network models, increasing the layers and time steps will enhance the memory ability of the model. On the contrary, if add too many layers, the complexity of the model will also increase significantly. Therefore, this paper uses the layer-by-layer optimization method [37]. That is, on the basis of the trained model, add new levels and retrain, and then fine-tune the whole network. The first layer of the proposed model MGRU takes geometric features as inputs, and the upper GRU layer takes the output from the lower GRU layer as the inputs. After three layers of the GRU model training, the final results of fault diagnosis are obtained.

C. OVERVIEW OF OUR APPROACH ARCHITECTURE

The process of the fault diagnosis is shown in FIGURE 3. In the proposed framework, the experimental steps of fault diagnosis are summarized as follows:

1) Step 1: Collect vibration signals from sensors installed on test-bench;
2) Step 2: Features calculating. Extract the time domain $\{A_i | i = 1, 2, \ldots 12\}$ and the time-frequency domain $\{B_i | i = 1, 2, \ldots 8\}$ of the vibration signals.
3) Step 3: Determine parameter values *layer* and *epoch*. Classification of each spur gear status $\{s\_i | i = 1, 2, \ldots 10\}$ by executing MGRU model.
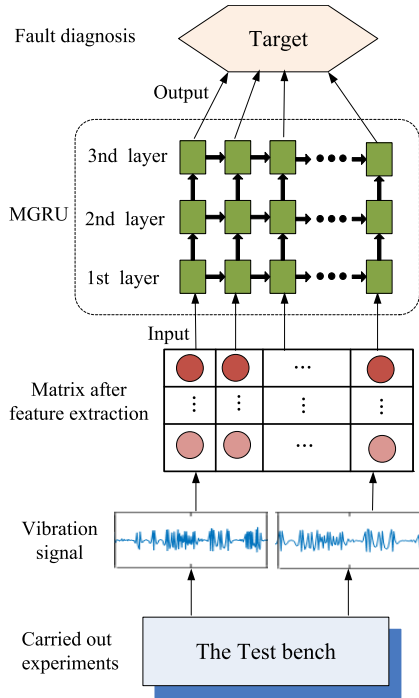
FIGURE 3. The architecture of our proposed approach. This framework is able to classify target from vibration signal.
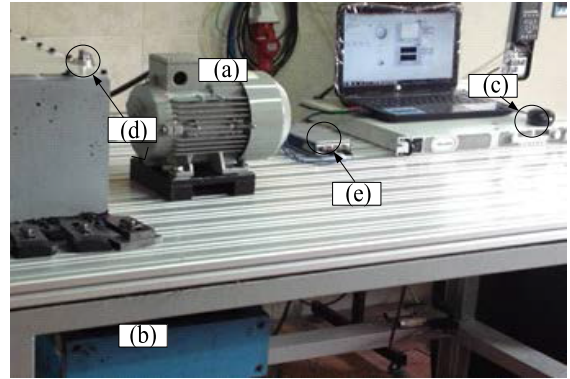

FIGURE 4. Experimental configurations: E xperimental set-up. (a): Motor; (b): Electromagnetic torque break, (c): Controller; (d): Accelerometer; (e): Data acquisition box.

D. PRECISION INDEX

To assess the accuracy, the criterion is formulated as follows:

$$Acc = \frac{S_r}{S_t} \times 100\% \qquad (15)$$

where S_r is the right number and S_t is the total number.

E. BASELINE

To evaluate the method's performance, four methods were applied to compare, i.e., GRU, LSTM, MLSTM, and support vector machine (SVM), respectively. The GRU model and LSTM model have been introduced in detail in Section II. In order to maintain the consistency principle of experimental comparison, the structure of the MLSTM is similar to MGRU to a great extent. But the units with three gates are used in MLSTM.

The SVM is a typical machine learning algorithm [38], and it is widely used in the field of fault diagnosis, so this paper chooses it as a comparative model. According to the limited sample information in machine statistical learning, it is able to find the best balance between the complexity of the model and the learning ability. It can map samples from original space to high-dimensional feature space, and separates them as far as possible in order to obtain the best classification ability. Given a training set $\{(x_i, y_i)|i = 1, 2, \ldots n\}$ with the input vector $x_i \in R^n$ and the output value $y_i \in \{-1, 1\}$, n is the number of training samples. They can be separated by a hyper plane, it is as follows:

$$wx + b = 0 \qquad (16)$$

where w is the weight coefficient vector in the feature space, and $b \in R$ is a constant. For linear inseparable problems, the kernel function k is needed to achieve linear separable.

$$\varphi(a) = \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} a_i a_j y_i y_j k\left(x_i, x_{j,}\right)$$

$$s.t. \sum_{i=1}^{l} a_i y_i = 0, \quad 0 \le a_i \le B \qquad (17)$$

where $k(x_i, x_j)$ is the kernel function, a_i is the Lagrange operator; B is the punishment factor, that is, tolerance of errors. Generally, the radial basis function (RBF) is used as the kernel function [36]. It can be expressed as follows:

$$k(x_i, x_{j,}) = \exp(-g \left\| x_i - x_j \right\|^2) \quad g > 0 \qquad (18)$$

where g is the kernel parameter, it implicitly determines the distribution of data after mapping to a new feature space.

III. EXPERIMENTAL TEST BENCH

In this study, experiments are carried out on the gear box fault diagnosis test bench. The experimental system is shown in FIGURE 4. The coupling connects the motor (Specification: SIEMENS, 3~, 2.0HP), the input shaft of the gearbox. The output shaft of the gearbox is connected with the electromagnetic torque circuit breaker (Specification: ROSATI, 8.83kW). The controller (Specification: TDK-Lambda, GEN 100-15-IS510) manually adjusts the torque breaking load. The vibration signal is collected by an accelerometer (Specification: PCB ICP 353C03), and then through a data acquisition box (Specification: DAQ, NI cDAQ-9234), it is sent to a laptop computer.

As shown in FIGURE 5, two spur gears (number of teeth Gear1 = 53, and Gear2 = 80) were installed on the gearbox. Fault diagrams and type descriptions of single spur gears are shown in TABLE 2. There are nine failure modes in combination of the two spur gears, plus a healthy state, there are ten different modes. The type state diagram of spur gear faults is indicated in TABLE 3.

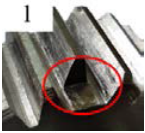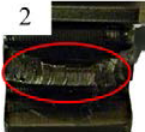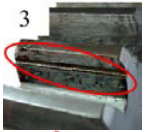**TABLE 2.** Condition patterns of the sun gear used in the experiment.
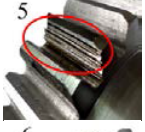
Fault State Diagram	fault type	Description
1	Broken tooth 25%	1: a small part of the gear teeth has been broken.
2	Broken tooth 50%	2: half of the teeth of the gear are broken.
3	Chaffing tooth	3: The whole independent groove appears.
4	Missing tooth	4: breaking of whole teeth.
5	Worn tooth	5: cracks like meshes.
6	Crack tooth 25%	6: the slippery part of the tooth surface produces shallow grooves.
7	Crack tooth 100%	7: the sliding part of the tooth surface produces deeper grooves.
8	Chaffing tooth 50 %	8: half independent groove appears.

TABLE 3. The type of fault in the spur gear.

state modes	Gear1	Gear2
s_1	Normal	Normal
s_2	Chaffing tooth	Normal
s_3	Worn tooth	Normal
s_4	broken tooth 25%	Normal
s_5	broken tooth 50%	Normal
s_6	Missing tooth	Normal
s_7	Normal	Crack tooth 25%
s_8	Normal	Crack tooth 100%
s_9	Normal	Chaffing tooth 50 %
s_10	broken tooth 25%	Crack tooth 25%

In the experiment, the load condition is 30V. With the sampling frequency 50 kHz, signals each with the duration 0.5 sec were collected. After receives the vibration signal,
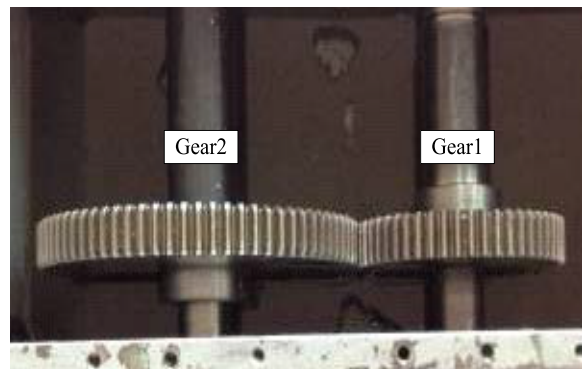

FIGURE 5. The spur gear configuration.

4500 sample points are consist the whole data sets. Each fault condition includes 5000 sample points. In the experimental process, the sample is split into two sets: 3000 samples for training and 1500 samples for testing. Since there are 20 features, a total of 4500 characteristic signals in the time domain and time-frequency domain are extracted. Then a 4500×20 original feature matrix is obtained.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In the task of gearbox fault diagnosis, as mentioned before, the 3000 samples are chosen to train the classifiers, and the 1500 samples are selected for testing. For each classifier, five experiments were carried out to compare the stability of the classification results. The training and testing data are selected randomly from the compressed feature matrix in each trial.

In order to test the performance of the designed MGRU spur gear fault diagnosis, three capabilities of accuracy, separability and robustness are studied respectively.

A. CLASSIFICATION CAPACITY

Through the previous work [27] and the characteristics of our data, structures of MGRU are defined as that the number of hidden layers are 3, 400 nodes in the first hidden layer, 300 nodes in the second hidden layer, and 200 nodes in the third hidden layer, respectively. The epoch number is 50. Based on these settings, the average result is 98.13%.

Regarding the MLSTM model [37], the nodes of layer 1, layer 2 and layer 3 are set to be 200, 100 and 100, respectively. The epoch number is also 50. In this case, the average of accuracy is 94.08%.

Regarding the GRU and the LSTM, the hidden layer units are 300 and 100 respectively. The epoch number is 50, too. The average of the five experiments results is 94.41% in GRU and 93.62% in LSTM respectively.

Regarding the SVM model in this paper, B is 1.0 and g is 200, and the average result is 91.4%.

Detailed results are shown in TABLE 4, and the Box-plot in FIGURE 6. Four findings can be drawn:

1) Among the MGRU, the GRU, the MLSTM, the LSTM and the SVM models, the MGRU has the highest average classification accuracy in fault diagnosis.

TABLE 4. Five test results of different models and their average accuracy.

Model	Value of *Acc* in each time (%)					Average
	1	2	3	4	5	
MGRU	97.61	98.26	97.86	98.53	98.4	98.13
GRU	95.66	94.12	93.78	94.36	94.12	94.41
MLSTM	95.13	95.33	91.60	95.86	96.06	94.08
LSTM	93.92	92.07	94.64	93.4	94.07	93.62
SVM	91.46	90.83	91.3	91.02	92.38	91.40
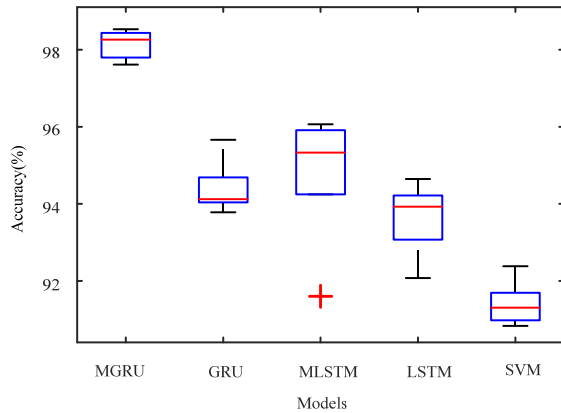
FIGURE 6. Box-plot of different models.

FIGURE 7. 2D Projection of Fault classification results. The representation is learned by MGRU.

TABLE 5. Settings with different parameters.

Experi-ment	*Layer_1*	*Layer_2*	*Layer_3*	*epoch*
ex_1	[600,500,400 ,300,200]	300	200	50
ex_2	400	[500,400,300, 200,100]	200	50
ex_3	400	300	[400,300,200, 100,50]	50
ex_4	400	300	200	[70,60,50 ,40,30]

2) Multi-layer neural network has better classification effect than single-layer neural network. The MGRU model is improved by 3.72% on the basis of GRU and MLSTM improved by 0.46% on the basis of LSTM.
3) The four neural network models (MGRU, GRU, MLSTM, and LSTM) perform better than the traditional data-driven model of SVM.
4) When these five experiments were carried out, the value distribution of MGRU model was concentrated and the veracity is highest. The MLSTM experiment produces outliers. The other three models (GRU, LSTM, and SVM) deviate greatly from the maximum and minimum values.

B. SEPARABILITY

Mechanical signals are chaotic. Whether the proposed model can extract effective information from them and classify them accurately is very important. To illustrate the effectiveness of the proposed method intuitively, the representation of fault classification results are visualized in FIGURE 7. In this paper, principal components analysis (PCA) algorithm [39]–[41] are used to reduce the dimension of features and project the high-dimensional representation into two-dimensional space. It can compress data as much as possible to retain the main features of variables, while minimizing the loss of information. Scatters of different colors and shapes represent ten different categories from s_1 to s_10 in Table 3, respectively.
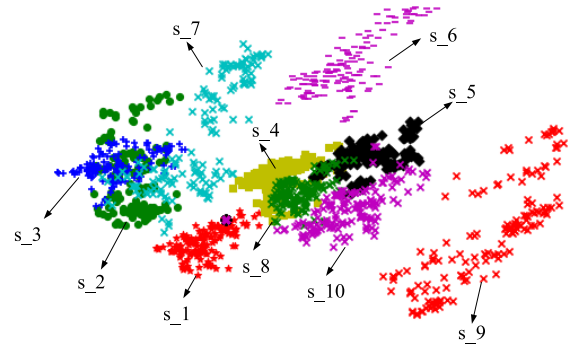
The graph shows that there are a small part of the classification results of s_2 and s_7, s_4 and s_8 coincide, because of signal aliasing when collecting vibration signals. It may exist due to some uncontrollable factors. However, several other fault types show a strong classification effect, especially the s_1, s_3, s_5, s_6, s_9, s_10 fault types.

The separability indicates that our model can learn from mechanical signals to identify fault types, and through training, can carry out strong information representation, which provides a powerful tool for fault diagnosis.

C. ROBUSTNESS

In addition, the robustness of our approach was investigated. There have two main parameters in the MGRU model, that is, the number of neurons per layer, *Layer_i*, (*i* = 1, 2, 3) and the number of iterations *epoch*. Four experiments are shown in the TABLE 5 for testing its robustness. In the ex_1, we vary the number of *Layer_1* in the range [600, 500, 400, 300, 200], and keep other parameters unchanged as original parameters (The original parameters are: *Layer_1* = *400*, *Layer_2* = *300*, *Layer_3* = *200* and the *epoch* = 50). In the same way, compared with the original parameters, the different in ex_2 just changed the number of *Layer_2* with the range [500, 400, 300, 200, 100], the number of *Layer_3* in the range [400, 300, 200, 100, 50] is changed in ex_3, and the range of *epoch* is [70, 60, 50, 40, 30] in ex_4.
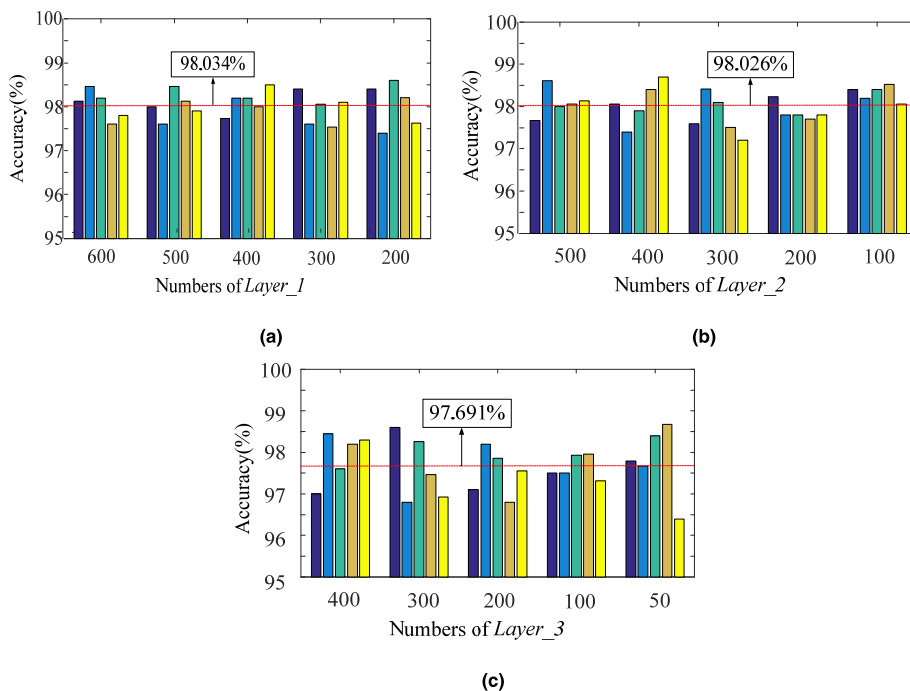
FIGURE 8. Performances of MGRU under different parameters of the sizes of different layers. (a) is the results of ex_1; (b) is the results of ex_2; (c) is the results of ex_3.

FIGURE.8 summarizes the classification accuracy using different numbers of the layer units. The same experiment was repeated five times for each combination of parameters. The abscissa represents the parameter values of the experimental robustness, and the ordinate represents the classification accuracy. The experimental results of ex_1 are shown in FIGURE. 8(a). In ex_1, When the parameter set is (*Layer_1 = 300, Layer_2= 300, Layer_3 = 200* and the *epoch = 50*), the average prediction accuracy of the five experiments is the worst, 97.938%; when the parameter set is (*Layer_1= 400, Layer_2 = 300, Layer_3 = 200* and the *epoch = 50*), the average prediction accuracy of the five experiments is the best, 98.126%. So even though the parameters vary greatly, the effect on the experimental results is only 0.188%. The experimental results of ex_2 are shown in FIGURE.8(b). The best average prediction accuracy of five experiments is 98.318%, the worst is 97.764%, and the difference is 0.554%. Similarly, the experimental results of ex_3 are shown in FIGURE.8(c). The best average prediction accuracy of five experiments is 97.912%, the worst is 97.504%, and the difference is 0.408%.

FIGURE.9 shows the classification accuracy using different numbers of the epoch numbers. The best average prediction accuracy of five experiments is 98.01%, the worst is 97.222%, and the difference is 0.788%.

The red lines in FIGURE.8 and FIGURE.9 show the overall average accuracy of the experiment in ex_1, ex_2, ex_3 and ex_4. They are 98.034%, 98.026%, 97.691% and 97.566% respectively. In summary, the different setting of parameters has little effect on classification accuracy.
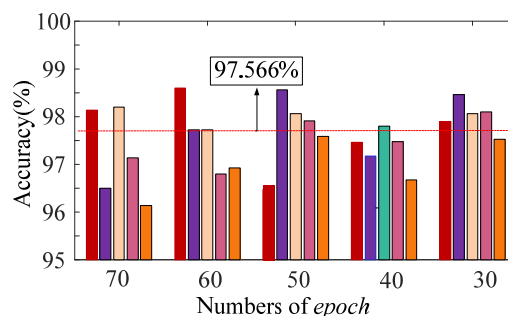

FIGURE 9. Performances of MGRU under different epoch numbers, that is, ex_4.

Obviously, the robust performance of the proposed method validates its generalization ability to diagnose various fault severity and mechanical equipment faults under different parameters.

V. CONCLUSION
Fault diagnosis plays an essential role in equipment maintenance. This paper introduces a fault diagnosis method based on a multi-layer GRU model. In this method, a deeper feature simulation is performed by adding layers of neural networks. Specifically, the vibration signals of healthy spur gears and nine different fault types are firstly collected on the experimental platform. Secondly twelve kinds of time domain feature extraction and eight WPT time-frequency domain feature extraction are carried out, and they are input into the three-layer GRU model to diagnose fault at last.

To verify the classification accuracy, four comparative experiments is carried out, i.e., the LSTM, the MLSTM, the GRU, and the SVM. In addition, the diagnostic performance of the proposed method is verified in terms of separability and robustness analysis. The following conclusions are drawn:

1) Based on the average classification accuracy generated by five classifiers, the results show that the fault classification accuracy of multi-layer model is higher than that of single-layer model, and the proposed model MGRU has the best classification accuracy.

2) The separability analysis of the proposed method indicates that the MGRU model can extract the information behind the chaotic mechanical signals and separate the fault types. It has strong information representation ability.

3) Based on the robustness analysis of the different parameters in the experiment MGRU, the fluctuation of classification accuracy in each trial are relatively small, illustrating that the model is insensitive to parameter change and has high stability. Therefore, the proposed model has a great contribution to the research of fault diagnosis.

ACKNOWLEDGMENT

REFERENCES
[1] 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. doi: 10.1016/j.ymssp.2015.08.030.

[2] C. Li, J. L. V. de Oliveira, M. L. Cerrada, D. Cabrera, R. Sanchez, and G. Zurita, "A systematic review of fuzzy formalisms for bearing fault diagnosis," *IEEE Trans. Fuzzy Syst.*, to be published. doi: 10.1109/TFUZZ.2018.2878200.

[3] B. Chen, Z. Zhang, C. Sun, B. Li, Y. Zi, and Z. He, "Fault feature extraction of gearbox by using overcomplete rational dilation discrete wavelet transform on signals measured from vibration sensors," *Mech. Syst. Signal Process.*, vol. 33, pp. 275–298, Nov. 2012. doi: 10.1016/j.ymssp.2012.07.007.

[4] Y. Li, Y. Wei, K. Feng, X. Wang, and Z. Liu, "Fault diagnosis of rolling bearing under speed fluctuation condition based on Vold-Kalman filter and RCMFE," *IEEE Access*, vol. 6, pp. 37349–37360, Jul. 2018. doi: 10.1109/ACCESS.2018.2851966.

[5] C. Li *et al.*, "A comparison of fuzzy clustering algorithms for bearing fault diagnosis," *J. Intell. Fuzzy Syst.*, vol. 34, no. 6, pp. 3565–3580, 2018. doi: 10.3233/JIFS-169534.

[6] E. Scheurwegs, M. Sushil, S. Tulkens, W. Daelemans, and K. Luyckx, "Counting trees in random forests: Predicting symptom severity in psychiatric intake reports," *J. Biomed. Inform.*, vol. 75, pp. S112–S119, Nov. 2017. doi: 10.1016/j.jbi.2017.06.007.

[7] B.-S. Yang, X. Di, and T. Han, "Random forests classifier for machine fault diagnosis," *J. Mech. Sci. Technol.*, vol. 22, no. 9, pp. 1716–1725, Sep. 2008. doi: 10.1007/s12206-008-0603-6.

[8] Y. Bai *et al.*, "A comparison of dimension reduction techniques for support vector machine modeling of multi-parameter manufacturing quality prediction," *J. Intell. Manuf.*, pp. 1–12, Jan. 2018. doi: 10.1007/s10845-017-1388-1.

[9] H. Darong, K. Lanyan, C. Xiaoyan, Z. Ling, and M. Bo, "Fault diagnosis for the motor drive system of urban transit based on improved hidden Markov model," *Microelectron. Rel.*, vol. 82, pp. 179–189, Mar. 2018. doi: 10.1016/j.microrel.2018.01.017.

[10] G. Zhang, T. Yi, T. Zhang, and L. Cao, "A multiscale noise tuning stochastic resonance for fault diagnosis in rolling element bearings," *Chin. J. Phys.*, vol. 56, pp. 145–157, Feb. 2018. doi: 10.1016/j.cjph.2017.11.013.

[11] Y. Liu and Z. Ge, "Weighted random forests for fault classification in industrial processes with hierarchical clustering model selection," *J. Process Control*, vol. 64, pp. 62–70, Apr. 2018. doi: 10.1016/j.jprocont.2018.02.005.

[12] Y. Li, M. Xu, X. Liang, and W. Huang, "Application of bandwidth EMD and adaptive multiscale morphology analysis for incipient fault diagnosis of rolling bearings," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6506–6517, Aug. 2017. doi: 10.1109/TIE.2017.2650873.

[13] H. Yi, C. Y. Du, C. B. Li, A. G. Wu, and X. Ying, "Sensor fault diagnosis of superconducting fault current limiter with saturated iron core based on SVM," *IEEE Trans. Appl. Supercond.*, vol. 24, no. 5, Oct. 2014, Art. no. 5602805. doi: 10.1109/TASC.2014.2352391.

[14] G. Sun, Q. Hu, Q. Zhang, A. Qin, and L. Shao, "Fault diagnosis for rotating machinery based on artificial immune algorithm and evidence theory," in *Proc. 27th Chin. Control Decis. Conf. (CCDC)*, May 2015, pp. 2975–2979. doi: 10.1109/CCDC.2015.7162380.

[15] F. Ye, Z. Zhang, K. Chakrabarty, and X. Gu, "Board-level functional fault diagnosis using artificial neural networks, support-vector machines, and weighted-majority voting," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 32, no. 5, pp. 723–736, May 2013. doi: 10.1109/tcad.2012.2234827.

[16] J. Ni, C. Zhang, and S. X. Yang, "An adaptive approach based on KPCA and SVM for real-time fault diagnosis of HVCBs," *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1960–1971, Jul. 2011. doi: 10.1109/TPWRD.2011.2136441.

[17] C. Li, Y. Tao, W. Ao, S. Yang, and Y. Bai, "Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition," *Energy*, vol. 165, pp. 1220–1227, Dec. 2018. doi: 10.1016/j.energy.2018.10.113.

[18] S. Zhang, Z. Sun, J. Long, C. Li, and Y, Bai, "Dynamic condition monitoring for 3D printers by using error fusion of multiple sparse auto-encoders," *Comput. Ind.*, vol. 105, pp. 164–176, Feb. 2019. doi: 10.1016/j.compind.2018.12.004.

[19] Y. Bai, Y. Li, B. Zeng, C. Li, and J. Zhang, "Hourly $PM_{2.5}$ concentration forecast using stacked autoencoder model with emphasis on seasonality," *J. Cleaner Prod.*, vol. 224, pp. 739–750, Jul. 2019. doi: 10.1016/j.jclepro.2019.03.253.

[20] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296–2305, Mar. 2017. doi: 10.1109/TIE.2016.2627020.

[21] V. N. Ghate and S. V. Dudul, "Cascade neural-network-based fault classifier for three-phase induction motor," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1555–1563, May 2011. doi: 10.1109/TIE.2010.2053337.

[22] X. Li, F. Duan, P. Loukopoulos, I. Bennett, and D. Mba, "Canonical variable analysis and long short-term memory for fault diagnosis and performance estimation of a centrifugal compressor," *Control Eng. Pract.*, vol. 72, pp. 177–191, Mar. 2018. doi: 10.1016/j.conengprac.2017.12.006.

[23] K. Cho *et al.* (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation." [Online]. Available: https://arxiv.org/abs/1406.1078

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

[25] M. Jung, H. Lee, and J. Tani, "Adaptive detrending to accelerate convolutional gated recurrent unit training for contextual video recognition," *Neural Netw.*, vol. 105, pp. 356–370, Sep. 2018. doi: 10.1016/j.neunet.2018.05.009.

[26] Y. Bai, B. Zeng, C. Li, and J. Zhang, "An ensemble long short-term memory neural network for hourly $PM_{2.5}$ concentration forecasting," *Chemosphere*, vol. 222, pp. 286–294, May 2019. doi: 10.1016/j.chemosphere.2019.01.121.

[27] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018. doi: 10.1109/TIE.2017.2733438.

[28] A. Qin, Q. Hu, Y. Lv, and Q. Zhang, "Concurrent fault diagnosis based on Bayesian discriminating analysis and time series analysis with dimensionless parameters," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2254–2265, Mar. 2019. doi: 10.1109/jsen.2018.2885377.

[29] Z. Li *et al.*, "Online implementation of SVM based fault diagnosis strategy for PEMFC systems," *Appl. Energy*, vol. 164, no. 15, pp. 284–293, Feb. 2016. doi: 10.1016/j.apenergy.2015.11.060.

[30] Q. Hu, A. Qin, Q. Zhang, J. He, and G. Sun, "Fault diagnosis based on weighted extreme learning machine with wavelet packet decomposition and KPCA," *IEEE Sensors J.*, vol. 18, no. 20, pp. 8472–8483, Oct. 2018. doi: 10.1109/JSEN.2018.2866708.

[31] J. Lin, "Feature extraction of machine sound using wavelet and its application in fault diagnosis," *NDT E Int.*, vol. 34, no. 1, pp. 25–30, Jan. 2001. doi: 10.1016/S0963-8695(00)00025-6.

[32] H. Yan, Y. Xu, F. Cai, H. Zhang, W. Zhao, and C. Gerada, "PWM-VSI fault diagnosis for a PMSM drive based on the fuzzy logic approach," *IEEE Trans. Power. Elect.*, vol. 34, no. 1, pp. 759–768, Jun. 2019. doi: 10.1109/TPEL.2018.2814615.

[33] I. Morgan and H. Liu, "Predicting future states withn-dimensional Markov chains for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 56, no. 5, pp. 1774–1781, May 2009. doi: 10.1109/TIE.2008.2011306.

[34] C. Li, R. V. Sánchez, G. Zurita, M. Cerrada, and D. Cabrera, "Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning," *Sensors*, vol. 16, no. 6, p. 895, Jun. 2016. doi: 10.3390/s16060895.

[35] R. Duan and F. Wang, "Fault diagnosis of on-load tap-changer in converter transformer based on time–frequency vibration analysis," *IEEE Trans. Ind. Electron*, vol. 63, no. 6, pp. 3815–3823, Jun. 2016. doi: 10.1109/TIE.2016.2524399.

[36] C. C. Chang and Z. Sun, "Structural damage assessment based on wavelet packet transform," *J. Struct. Eng.*, vol. 128, no. 10, pp. 1354–1361, Oct. 2002. doi: 10.1061/(ASCE)0733-9445(2002)128:10(1354).

[37] S. Zhang, X. Liu, and J. Xiao, "On geometric features for skeleton-based action recognition using multilayer LSTM networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 148–157. doi: 10.1109/WACV.2017.24.

[38] Y. Xiao, Y. Hong, X. Chen, and W. Chen, "The application of dual-tree complex wavelet transform (DTCWT) energy entropy in misalignment fault diagnosis of doubly-fed wind turbine (DFWT)," *Entropy*, vol. 19, no. 11, p. 587, Jan. 2017. doi: 10.3390/e19110587.

[39] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 32–55, Jul. 2018. doi: 10.1109/MSP.2018.2826566.

[40] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015. doi: 10.1016/j.neunet.2014.09.003.

[41] J. Dozier, T. H. Painter, K. Rittger, and J. E. Frew, "Time–space continuity of daily maps of fractional snow cover and albedo from MODIS," *Adv. Water Resour.*, vol. 31, no. 11, pp. 1515–1526, Nov. 2008. doi: 10.1016/j.advwatres.2008.08.011.
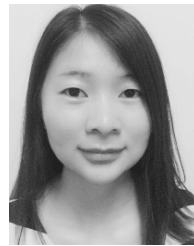
XIAODAN WANG was born in Shangqiu, Henan, China, in 1988. She received the B.S. degree in management from the Henan University of Chinese Medicine, China, in 2017. Since 2017, she has been a graduate with the National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing, China, majoring in management science and engineering. Her research interests include fault diagnosis, risk prediction, and health management of manufacturing systems.

RENÉ-VINICIO SÁNCHEZ received the B.S. degree in mechanical engineering from the Universidad Politécnica Salesiana (UPS), Ecuador, in 2004, and the Ph.D. degree in industrial technologies research from the Universidad Nacional de Educación a Distancia (UNED), Spain, in 2017. He is currently a Professor with the Department of Mechanical Engineering, UPS. His research interests are in machinery health maintenance, pneumatic and hydraulic systems, artificial intelligence, and engineering education.

SHUAI YANG (M'17) was born in Chongqing, China, in 1986. He received the B.S. and M.S. degrees in mechanical engineering from Chongqing University, China, in 2012, and the Ph.D. degree in mechanical engineering from the University of Ottawa, ON, Canada, in 2017.

He is currently an Associate Professor and a Research Assistant with the National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing. He is the author of a book, over 15 articles, and over 5 inventions. His research interests include vibration control, fault diagnosis, and health management of the manufacturing systems. He was a recipient of the Excellent Reward of graduate studies from the Canadian Society of Engineering, in 2016.

YING TAO (M'17) was born in Xinyang, Henan, China, in 1996. She received the B.S. degree in management from the Henan University of Chinese Medicine, China, in 2017. She is currently pursuing the M.S. degree with the Chongqing Technology and Business University, China. She is also a graduate with the National Research Base of Intelligent Manufacturing Service, majoring in management science and engineering. Her research interests are in data analysis and machinery health maintenance.

YUN BAI (M'17) received the Ph.D. degree from Chongqing University, China, in 2014. He is currently an Assistant Professor with the National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing, China, and is also a Postdoctoral Researcher with the University of Algarve, Portugal. His current research interests include intelligent system modeling and management.

• • •