Research Article

Multilayer Extreme Learning Convolutional Feature Neural Network Model for the Weak Feature Classification and Status Identification of Planetary Bearing

Rui Yang [ID], Zenghui An, and Shijun Song

School of Mechanical and Electronic Engineering, Shandong Jianzhu University, Jinan 250101, China

Correspondence should be addressed to Rui Yang; woyangr@126.com

A convolutional neural network has the characteristics of sharing information between layers, which can realize high-dimensional data processing. In general, the convolutional neural network uses a feedback mechanism to realize parameter self-regulation, which solves the disadvantages of manual parameter adjustment. However, it is unable to determine the iteration number with the best calculation accuracy. Calculation efficiency cannot be guaranteed while achieving the best accuracy. In this paper, a multilayer extreme learning convolutional neural network model is proposed for feature recognition and classification. Firstly, two-dimensional spatial characteristics of planetary bearing status data were enhanced. Then, extreme learning machine is embedded in a convolution layer to solve convex optimization problems. Finally, the parameters obtained from the training model were nested into a network to initialize the model parameters to separate each status feature. Planetary bearing experimental cases show the effectiveness and superiority of the proposed model in the recognition and classification of weak signals.

1. Introduction

With the improvement of automation level in a modern production system, rotating machinery presents the development direction of high speed, high efficiency, and maximum economic benefit. However, a continuous production process makes the equipment run under heavy load for a long time, which will easily lead to accelerated fatigue of transmission parts. Furthermore, tight connections between devices make the health status of an individual component affect the efficiency and quality of the entire system. Once the transmission parts fail, it will lead to a series of chain reactions and even make the whole equipment or even the whole production line stop working. Therefore, reliable monitoring of transmission parts is crucial to maintain the whole safety production process.

In recent years, a planetary gearbox with dual rotor bearings has become the main transmission component due to its series of advantages with compact structure, large transmission ratio, light weight, and strong bearing capacity. It is widely used in automotive, wind power, aerospace, and other fields. Because the equipment works in complex environment for a long time, it is easy to cause accelerated fatigue of transmission parts [1]. For example, wind energy as green energy has promoted wind power generation to become one of the fastest growing branches of the current power generation field. Typically, a wind turbine consists of a planetary gear train (I level transmission) and two fixed-shaft gear trains (II level and III level transmission), as shown in Figure 1. Planetary gear trains are usually mounted at the low-speed end to withstand greater torque. In addition, wind turbines are usually located in a relatively wide-open area or offshore areas and often affected by irregular variable speed winds and the external ambient temperature that change with the season. Due to the complex working environment, the key components (gears and bearing) of the planetary gearbox are easily damaged. For example, the G52-850 wind turbine, consisting of Gamesa and Echesa speed-increasing gearboxes and INDER generators, showed abnormalities after 5-year work. Through
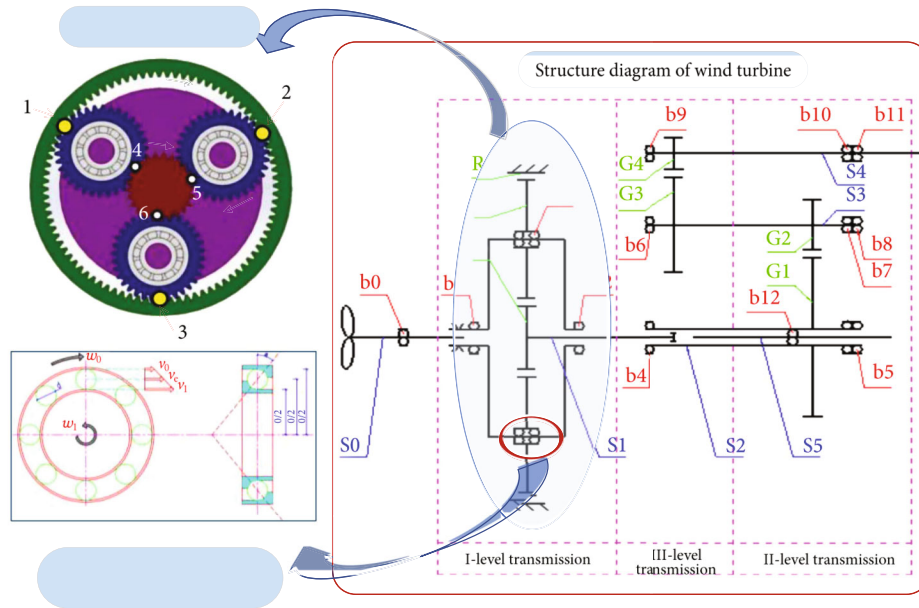
FIGURE 1: Layout of wind turbine transmission components and schematic diagram of planetary bearings.

endoscopic and unpacking tests, it was found that the fault was caused by planetary bearings [2]. It is the key to maintain the whole safety production process to adopt a reliable monitoring method to monitor the equipment condition. Therefore, the working efficiency and safety of a wind power generation system can be greatly improved when the planetary bearing is operated in stable status.

From the evolution process of bearing failure (Figure 2), it can be seen that the initial stage of failure accounts for a larger proportion of the entire damage. As the fault continues to deteriorate, the degradation rate increases exponentially. In the early stage of failure, the abnormal symptoms are slight, the impact on the mechanical system is small, and the maintenance cost is relatively low. If fault goes undiagnosed or unnoticed at an early stage, it will lead to a catastrophic accident when early fault develops and accumulates to a certain extent. Therefore, early detection, early diagnosis, and early maintenance are essential to ensure the safe operation of high-precision equipment. In addition, under the demand of intelligent devices, the amount of data that needs to be analyzed is also large. Traditional fault diagnosis based on point-by-point single-signal analysis is difficult to detect the characteristic components related to fault quickly and accurately, which seriously hinders the development process of high precision, high speed, and high reliability of high-end equipment, whereas intelligent fault diagnosis based on "data-driven" can solve this problem with high precision, high speed, and high reliability [3].

In 2006, the birth of the deep learning algorithm [4, 5] marks the development of fault diagnosis towards rapidity, efficiency, and intelligence. High target feature resolution of a data set will get accurate fault diagnosis results, and complete data volume can improve model learning ability. These existing data-driven neural network models have achieved good results in some ideal environments. However, there are still some factors restricting their application in the field of fault diagnosis. However, due to the randomness and uniqueness of faults, the field of intelligent diagnosis faces some bottlenecks. Rotating machinery is in a healthy status for a long time, and most of the collected signals are in a healthy status. Due to the high cost of collecting measured fault samples, it is difficult to obtain all types of fault samples, which makes the sample set unbalanced. Besides, in case of early fault or large external interference, the fault characteristic information is weak or even submerged. The model may give interference information a high confidence output. Aiming at the incomplete characteristics of the fault data set, Gao et al. and Liu et al. are committed to using finite element method simulation to simulate a sample with different fault statuses [6, 7]. An et al. [8] proposed a self-learning transferable neural network for fault intelligence diagnosis with unlabeled and imbalanced data. Most of the weak fault intelligent diagnosis methods [9, 10] use traditional fault feature extraction as the preprocessing to extract sensitive information, and there is a lack of research on improving the robustness of the model itself.

A convolutional neural network (CNN) [11, 12] is one of the representative models for intelligent recognition and classification of weak fault signals of bearings. It has attracted the attention of many researchers and been widely used in many fields such as bearing fault diagnosis. Fu et al. [13] used 1D convolution kernels of different scales to extract multiscale features and performed dimensional assimilation on feature space of different scales based on fusion theory to adapt to convolution operation. Zhao et al. [14] converted one-dimensional time-domain signal into 2D grayscale images, which were used as the analysis sample data of the CNN model. This solved the problem of insufficient data and avoided the process of artificial feature extraction. Cyclic spectral coherence was adopted as preprocessing to extract information that best characterized the status of bearing [15]. Then, group normalization
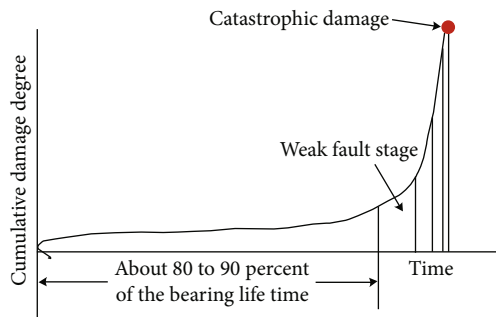
Figure 2: Failure development trend chart.

calculation was introduced to balance the distribution difference of data. Ye et al. [16] proposed a new method called deep morphological convolutional network, which consists of two parallel branches: noise filtering and feature selection algorithm. Noise filtering can update structure elements based on backpropagation. A feature selection algorithm was based on kurtosis weight fusion. Besides, values of various hyperparameters directly affect the training speed and accuracy of the CNN model. Currently, an error backpropagation (EBP) mechanism was often used to modify the model parameters. In the process of parameter adjustment, the initialization values of some parameters may also affect the classification results of the model. In addition, the range of adjustable parameters involved in the algorithm directly affects the computational complexity. Therefore, the CNN model is not suitable for rapid online monitoring, especially for early diagnosis of weak faults. An effective model is urgently needed to improve the performance of online monitoring.

In order to build a model framework with superior performance, the extreme learning machine (ELM) principle is adopted to deal with the convex optimization problem of a convolution layer. ELM was firstly proposed by Huang et al. [17] for a feedforward single-layer neural network. Subsequently, ELM was gradually introduced into the multilayer model structure [18, 19]. Compared with other models such as Deep Belief Network (DBN) [20, 21] and Stacked Autoencoder (SAE) [22], the ELM model involves fewer parameters and has higher computational efficiency and less complexity. Therefore, ELM has been favored by researchers in many fields, such as image processing [23], objective optimization [24, 25], dimensionality reduction [26], and fault diagnosis [27–29].

ELM was combined with other models to improve the training efficiency and recognition accuracy. For example, ELM was combined with an autoencoder to mine deep features of training data and proved to be superior to ELM, SAE, and CNN [30]. Online sequential ELM was proposed to classify and recognize the low-dimensional features extracted from the SAE model, whose effectiveness of this method has been proven for tool wear status recognition [31]. ELM was used as an enhanced classifier to improve the recognition accuracy of an integrating CNN model. Its superiority in training speed and accuracy was verified by comparing with other 6 models [32].

In general, ELM plays a role of an efficient classifier in the hybrid model. The existing model framework based on multilayer perceptron has shortcomings in improving the training speed. The goal of this paper presented here is to find a training model mechanism to improve the training accuracy and speed. Based on this, a fast and effective embedded hybrid model structure, called multilayer extreme learning convolutional feature neural network model (M_ELMConvNet), was proposed. The main contributions here are twofold.

(1) The wavelet cyclic spectrum feature extraction method [33] was used to convert the time-domain signal into a two-dimensional image. Then, the obtained image is partitioned, which is more suitable for CNN analysis

(2) A new model training mechanism of embedding ELM into a convolutional layer was proposed to improve the calculation speed and classification accuracy. The final classification and recognition results are obtained by multilayer stacking structure. The computational speed and accuracy of the proposed algorithm are verified by comparing with the results of other models

The remainder of this paper is structured as follows: the relevant theoretical research background contents, such as CNN and ELM models, are shown in Section 2. The proposed model framework and implementation process are introduced in Section 3. The proposed method is applied to the experimental data in Section 4. Finally, conclusions and the next step are described in Section 5.

2. Theoretical Background

2.1. CNN. CNN is a self-learning model that can automatically extract the internal feature information of the input data and implement classification tasks. Different from traditional neural networks, CNN generally contains a convolution layer and a subsampling layer (also called pooling layer). CNN learns hidden features by continuously running the convolutional layer in a loop and performing pooling operations. The convolution layer is used to convolute the original input data with multiple local filters to generate locally invariant feature information, which is used as the input of the pooling layer to extract representative features. The procedure is shown in Figure 3.

Suppose the size of the input layer is Height × Width = $a \times b$, and the number of input layers is c which is the number of channels. Define a convolution kernel as K, whose size is $n \times n$. The entire operation process of the convolution layer is to continuously perform convolution operations on the input layer data and the convolution kernel. The features extracted by the convolutional layer are served as the input of the pooling layer to further reduce the dimension of the feature matrix by calculating the local average or maximum. Subsequently, the fully connection layer tiles the output matrix of the pooling layer. The main task of each layer before the fully connected layer is feature extraction. The
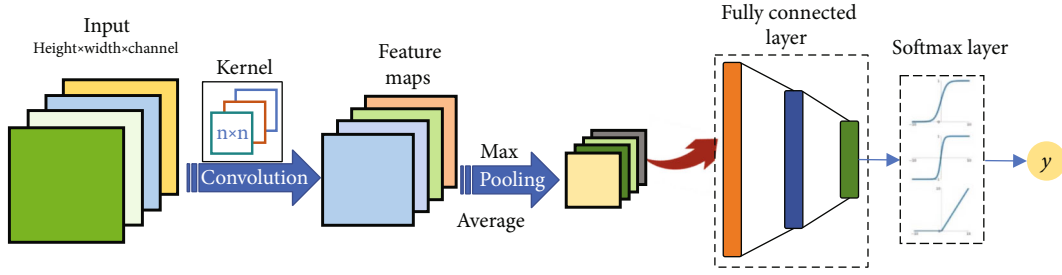
FIGURE 3: CNN model frame diagram.

classification task starts at the fully connection layer. Generally, there are multiple full connection layers in the whole network structure. Because the single-layer structure can only solve the linear classification problem and most of the problems in real life are nonlinear problems, the softmax layer is also connected behind the full connection layer to further predict the label. By calculating the probability of each sample's category, the label category with the largest probability value is assigned to the sample data.

2.2. ELM. ELM uses randomness and Moore-Penrose generalized inverse theory to calculate parameters, which avoids the use of EBP and greatly improves the training speed. The input layer data is usually nonlinear and separable. The core idea of the ELM algorithm is to map the original data into a high-dimensional feature space by adding hidden layer nodes on the premise that the input data is linearly separable. The connection weights between the input layer and the hidden layer are randomly generated. The connection matrix between the hidden layer and the output layer is calculated by Moore-Penrose generalized inverse. The entire training process only needs to adjust the number of hidden layer nodes. The schematic diagram is shown in Figure 4.

3. Proposed Architecture and Method

3.1. Formulated General Model Framework. Suppose there is a set of bearing status data to be classified. The data set are $D_x = \{x_1, \cdots, x_N\}, x_j \in R^{1 \times n \times c}$, and the target matrix is $D_Y = \{Y_1, \cdots, Y_N\}, Y_j \in R^C$, where n and c are the length and the number of channels of the input data, respectively. N is the number of samples in the data set. C is the number of target category to which the input data belongs. The relationship between the sample set and the label matrix can be achieved by the classification function, as shown in the following formula:

$$D_Y = f_c(D_x), \tag{1}$$

where f_c is the classification function of D_x.

3.2. Data Graphical Processing and Enhancement. In order to improve the accuracy of model recognition, the two-dimensional CNN model was used for data processing. The original sample data is one-dimensional, which reflects the time-domain waveform information and often overwhelms some fault active components. In this paper, one-

dimensional time-domain data are converted into 2D images based on wavelet cyclic spectrum theory [33] and the periodic characteristics of nonstationary bearing information are extracted. The local information in cycle spectrum of each fault type is similar. In order to make up for the shortage of faulty data samples and improve the quality of sample data, the converted image is further processed by block localization. Suppose that $W_s = \{S_1, \cdots, S_N\}, S_j \in R^{w \times h \times c}$ is the 2D cyclic spectrum sample matrix, w and h are the width and height of each single image sample matrix, respectively, and the label matrix is $W_L = \{L_1, \cdots, L_N\}, L_j \in R^L$. For the sample W_s, we regard it as being composed of several submodules W_P. $P_s = \{S_1^P, \cdots, S_N^P\} \subseteq W_P$, and corresponding labels $P_L = \{L_1^P, \cdots, L_N^P\}$.

$$P_L = f_p(P_I), \tag{2}$$

$$W_L = f_q\left(f_p \otimes W_P\right), \tag{3}$$

where $\otimes$ is the tensor product and f_p and f_q are the classification functions for P_I and W_s, respectively.

3.3. Parameter Transfer

3.3.1. Convolution Layer Detector Based on ELM. The ELM algorithm is integrated into the model to further improve the classification characteristics of the algorithm. As mentioned above, ELM is superior in complexity and accuracy compared with other algorithms [34]. The weight matrix of the input layer is randomized; that is, the weight matrix of the input layer and the input data are independent of each other. We can set it arbitrarily according to some distribution theories. The output matrix of the hidden layer and the weight matrix between the hidden layer and output layer need to be calculated relying on label data.

The specific implementation process is as follows:

Assuming there are N test samples (X_j, L_j), $X_j = \{x_{j,1}, x_{j,2}, x_{j,3}, \cdots x_{j,n}\}^T$ and corresponding labels are $L_j = \{l_{j,1}, l_{j,2}, l_{j,3}, \cdots l_{j,n}\}^T$. N_H is the number of hidden layer nodes. Hidden layer output can be expressed as

$$H_o = W_i \bullet X_j + b_i, \tag{4}$$

where $W_i = \{w_{i,1}, w_{i,2} \cdots w_{i,n}\}$ is the connection matrix between the input layer and hidden layer and b_i is the offset

(a) ELM network structure diagram and its parameterized model

(b) Spatial model diagram of ELM. The nonlinear characteristics of the input data are fitted by multiple mappings
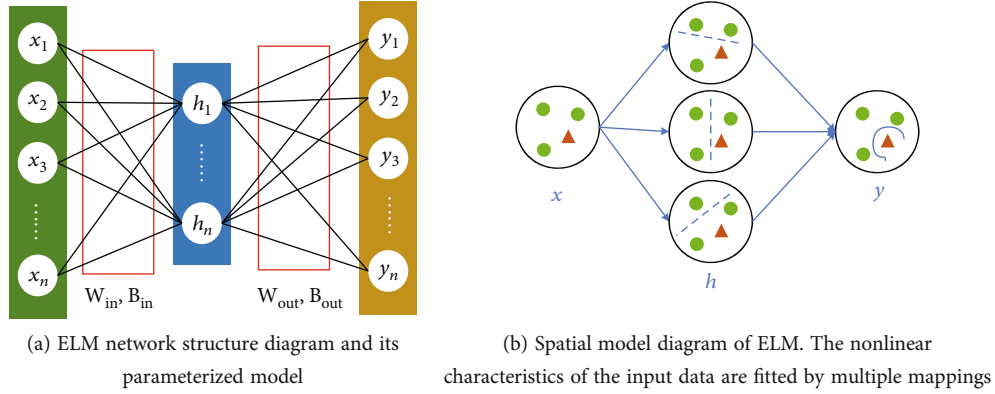
FIGURE 4: Schematic diagram of ELM structure and principle.

vector. Both of them are randomly generated. The output layer matrix can be expressed as

$$O_j = \begin{bmatrix} f(W_1 \bullet X_1 + b_1) & \cdots & f\left(W_{N_H} \bullet X_1 + b_{N_H}\right) \\ \vdots & & \\ f(W_1 \bullet X_n + b_1) & \cdots & f\left(W_{N_H} \bullet X_n + b_{N_H}\right) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{N_H} \end{bmatrix} = f(H_o)\beta_i,$$

$$f(H_o) = \begin{bmatrix} \sigma_{\mathrm{sig}}(W_1 \bullet X_1 + b_1) & \cdots & \sigma_{\mathrm{sig}}\left(W_{N_H} \bullet X_1 + b_{N_H}\right) \\ \vdots & & \\ \sigma_{\mathrm{sig}}(W_1 \bullet X_n + b_1) & \cdots & \sigma_{\mathrm{sig}}\left(W_{N_H} \bullet X_n + b_{N_H}\right) \end{bmatrix}. \quad (5)$$

$O_j = \{o_1, o_2, \cdots o_n\}$ indicate the categories of the output target label. β_i is the layer connection weight (LW) between the hidden layer and output layer. β_i is unknown and must be calculated based on the label data in the process of prediction classification results of the ELM algorithm. The expression is

$$\beta = H_o^{-1}L. \quad (6)$$

In the training stage, the number of hidden layer nodes is an uncertain factor, which has an impact on the prediction performance. In order to evaluate the performance of the parameters calculated by the model, the error minimization loss function was used as an evaluation index of prediction ability, shown as

$$E = \min \sum_{j=1}^{N} \|O_j - L_j\|, \quad j = 1, 2, 3 \cdots N. \quad (7)$$

The parameters corresponding to the minimum E value are the optimal values that best characterize the target features. Finding the minimum number of hidden layers while ensuring the highest accuracy is another factor to improve operation efficiency.

3.3.2. Pooling Layer. The pooling layer is actually a down-sampling layer, which is mainly used to extract local features and prevent overfitting. The procedure is as follows: first, define the size and step of the local pool module. Then, the local feature extraction method is determined. The most common method is to calculate the average or maximum value of each module. In this paper, the pooling layer is followed by the random parametric dimension reduction layer, as shown in equation (4). The pooling process adopts the method of calculating the average value as shown in

$$A(u_1, u_2) = \sum_{u_1=1}^{N_H} \sum_{u_2=1}^{N_H} \frac{1}{r \times r} \sum_{i=1}^{r} \sum_{j=1}^{r} O(i + (u_1 - 1) \times r, j + (u_2 - 1) \times r), \quad (8)$$

where r is the size of each step.

3.3.3. Normalization. After the pooling layer and before the ELM classification layer, a min-max standardization process was added to prevent the occurrence of gradient disappearance. The normalization result is obtained by

$$S_x = \frac{A_x - \min (A_x)}{\max (A_x) - \min (A_x)} \times (N_{\max} - N_{\min}) + N_{\min}, \quad (9)$$

where min and max are operators for calculating the maximum and minimum elements in the matrix A_x, respectively. $N_{\max}$ and $N_{\min}$ are the range of the interval for normalizing the matrix.

3.4. Test Model. Intelligent diagnosis methods based on deep learning theory mostly rely on a large amount of training data to achieve classification and recognition. The premise is that the performance of connection weights must be evaluated and modified on the basis of data with accurate label. In this paper, a method based on parameter transfer theory is proposed. The existing labeled data is input into the supervised model for training, and the connection weights of each module are obtained. Afterward, interlayer connection matrices are input into the test model to recognize and classify the unlabeled data.
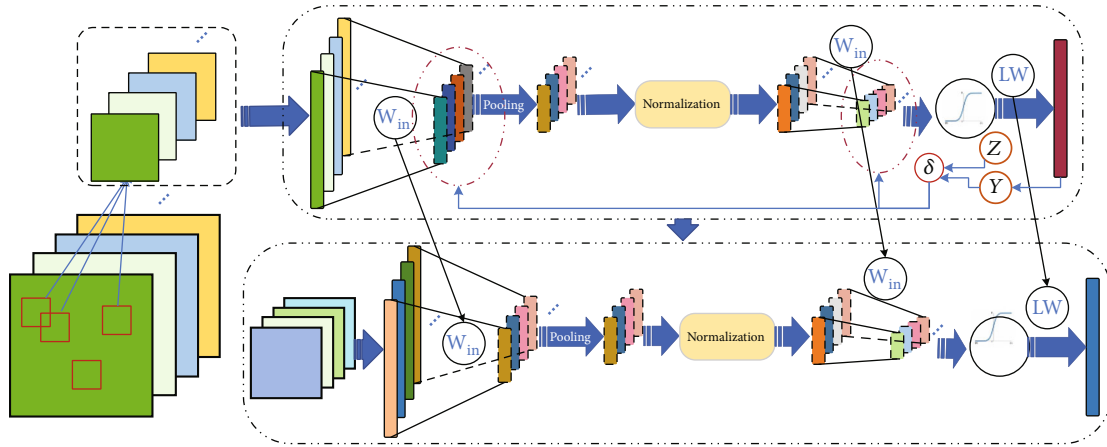
FIGURE 5: Framework of the proposed algorithm.

3.5. Fault Diagnosis Method Based on M_ELMConvNet. This paper presents a fast feature learning method based on two-dimensional CNN and ELM, and the model frame is shown in Figure 5.

Figure 6 shows the flowchart of the present method, and the procedures are as follows:

Firstly, samples were collected and processed with 2D data transformation and enhancement.

(1) The acceleration signals of bearings in four different working conditions were collected. About 85% of the data were labeled as training samples. The rest of the data is unmarked and is considered the test sample

(2) The original data is transformed into a 2D image by wavelet cyclic spectrum analysis

(3) Subsequently, the obtained image is partitioned to enhance the data according to equation (3)

Secondly, an efficient and accuracy neural network classification model is constructed.

(1) The entire model framework consists of two parts. One is the training process for labeled data, and the other is the predictive classification process for unlabeled data

(2) The labeled data is input into the supervised training model. The first step of this model is to reduce the dimension of the data by randomizing the connection matrix. The specific implementation process was based on the ELM training principle, as shown in equation (4). Then, the dimensionless data is entered into the pooling layer and standardized. Finally, the data is input to ELM for supervisory testing. The error rate between the test data and the label data is used as the loss function to adjust the random parameterized dimensionality reduction and the number of ELM hidden nodes. Subsequently, the optimal node number was assigned to the model to predict the connection matrix between each layer

(3) In the prediction process of the test sample set, based on the idea of parameter transfer, the connection matrix $W_{1\text{in}}$, $W_{2\text{in}}$ and layer connection weight LW obtained by the supervised test process were input into the test model of the corresponding prediction sample as the preset model parameter values, and the final prediction results were obtained

Thirdly, the whole training process is applied to the recognition and classification of bearing status data.

4. Experimental Validation

The data for verifying the effectiveness of the proposed algorithm were obtained from the comprehensive test bench for power transmission fault diagnosis. The sample data analyzed were measured under different conditions at different times.

4.1. Experiment Setup and Data Description. The power transmission system of the testbed consists of a planetary gearbox, a parallel shaft gearbox supported by rolling bearings or sleeve bearings, bearing load, and programmable magnetic brake, as shown in Figure 7. The testbed includes all the necessary powertrain configurations for studying gearbox dynamics and noise characteristics, health monitoring techniques based on vibration signal analysis, lubrication conditions, and wear particle analysis. The testbed has stable performance and can withstand strong load impact. There is enough space for the replacement and installation of gears and the installation of a monitoring device. Planetary gear systems, sun gears, planetary gears and gear rings, brackets, and bearings are easy to be disassembled.

The vibration signals under four statuses: no damage, first-stage planetary bearing outer ring failure, first-stage planetary bearing inner ring failure, and first-stage planetary bearing ball failure, were collected for analysis. In the experiment, the relevant parameters are set as follows: sampling frequency is 15360 Hz and motor speed is 2100 r/min. Multiple acceleration signals under different working conditions were collected. The data collected under each health condition were separated into 470 equal parts. Randomly select
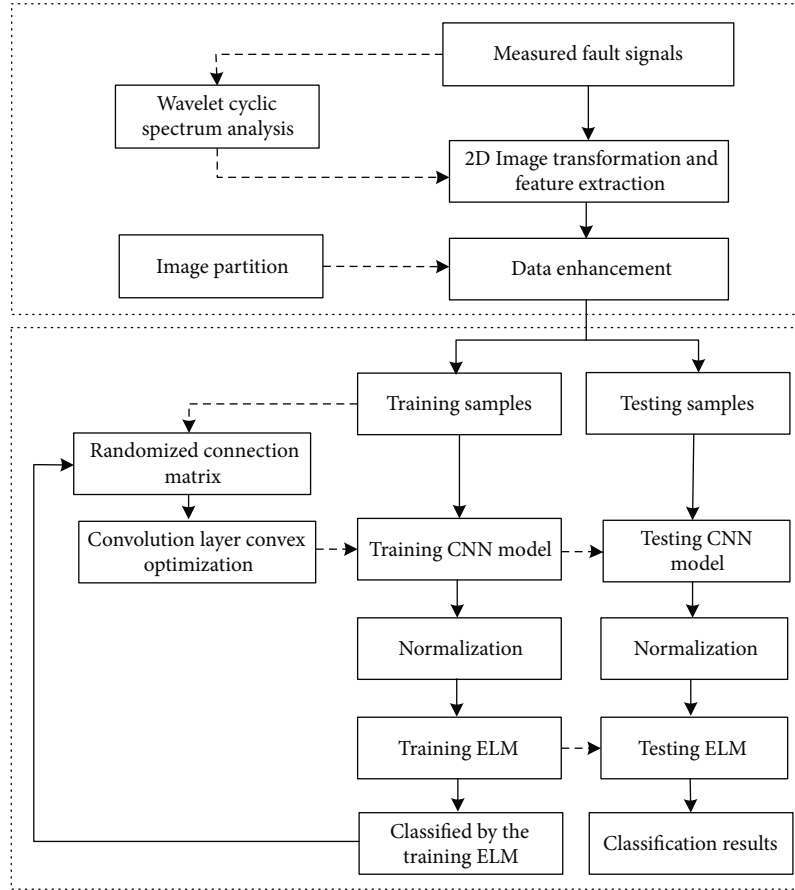
FIGURE 6: The flowchart of the proposed fault diagnosis method.

FIGURE 7: Comprehensive test bench for power transmission fault diagnosis.

TABLE 1: Description of status data sample set.

Planetary bearing damage statuses	No damage	Inner-race fault	Outer-race fault	Ball fault
Number of training samples	400	400	400	400
Number of test samples	70	70	70	70
Label	1	2	3	4

400 pieces of data in these four statuses as the training data set and 70 pieces as the test data set, respectively. A detailed partition of the sample set for status data analysis is shown in Table 1. The time waveforms for 2.5 s and the spectrum with bandwidth [0 Hz 2500 Hz] of the vibration signal for four statuses are shown in Figure 8. It can be seen from Figure 8 that the time-domain waveforms and spectrums for different four-status signals are inevitably affected by external interference information, which is also one of the factors that reduce the ability of model recognition.

4.2. Result Analysis. The purpose of this paper is to propose a fast and effective intelligent classification method for weak fault data of planetary bearings. In order to further verify the effectiveness of the proposed M_ELMConvNet neural network model, the experimental data were analyzed by the algorithm described in Section 3.5. For comparison, three other models including ELM, BP-based CNN, and Deep Autoencoder (DAE) were also applied to status identification of experimental data.

The average recognition accuracy of the algorithm under different hidden layer nodes is calculated by executing the model for 20 times. The results are shown in the box diagram in Figure 9. As shown in Figure 9, the prediction accuracy rate of most testing results was above 98%. The validity of M_ELMConvNet in planetary bearing status recognition is further verified. With the increase in the number of hidden layer nodes, the recognition accuracy rate fluctuates slightly. When the number of hidden layer nodes is set to 290, the average prediction accuracy is relatively high and
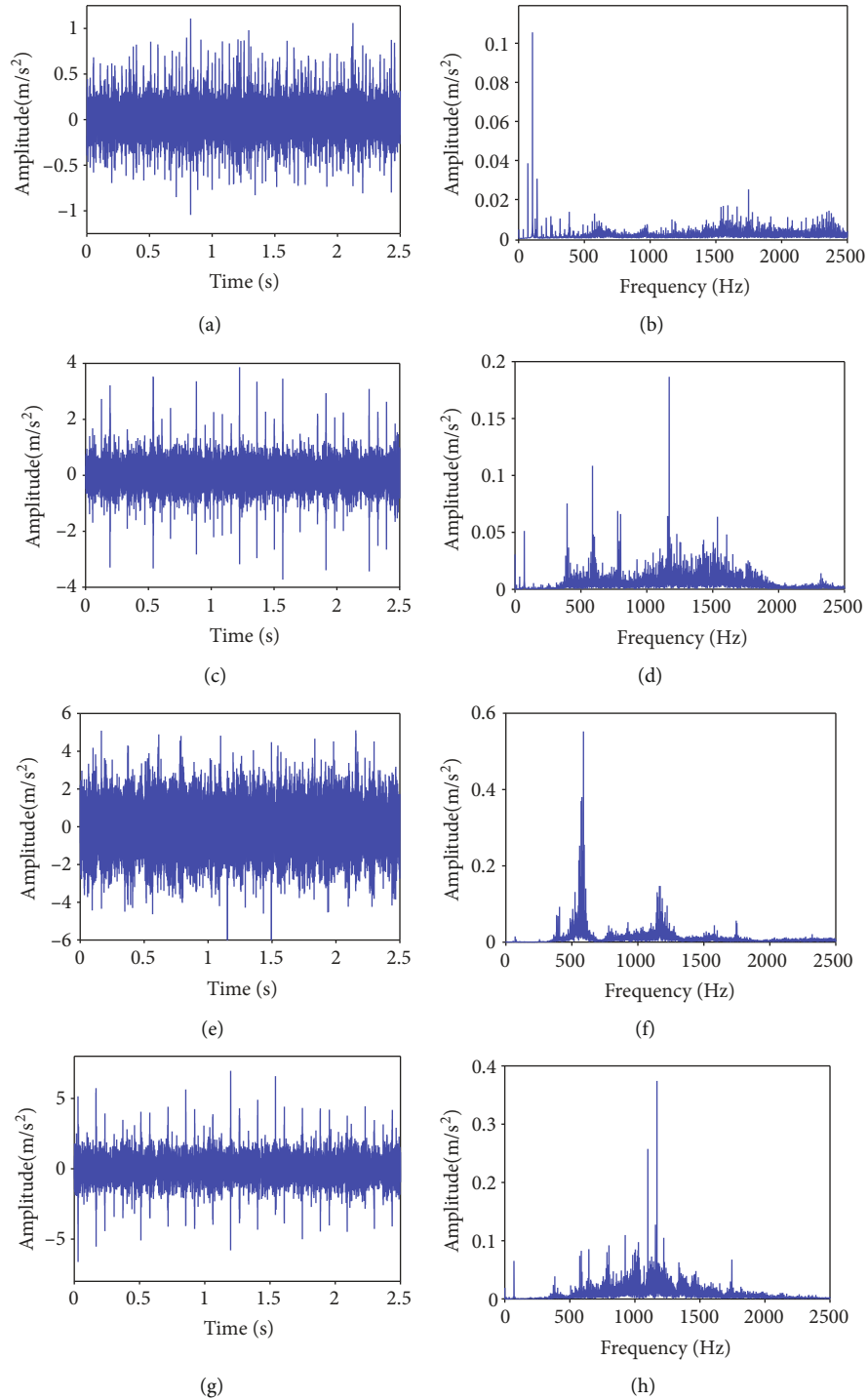
Figure 8: Signal waveforms listed in Table 1: (a) no-damage bearing signal in the time domain; (b) the spectrum of no-damage bearing signal; (c) inner-race fault bearing signal in the time domain; (d) the spectrum of inner-race fault bearing signal; (e) outer-race fault bearing signal in the time domain; (f) the spectrum of outer-race fault bearing signal; (g) ball fault bearing signal in the time domain; (h) the spectrum of ball fault bearing signal.

the stability is strong. Meantime, the model training time also showed an exponential growth trend, as shown in Figure 10. In the subsequent analysis, the number of hidden layer nodes was set to 290 based on the balance training time and prediction accuracy.

Figure 11 is the confusion matrix of multistatus classification and recognition accuracy based on the proposed method. As can be seen from Figure 11, the highest prediction accuracy is 100% for status 4. The minimum recognition accuracy is status 3 because there is no obvious
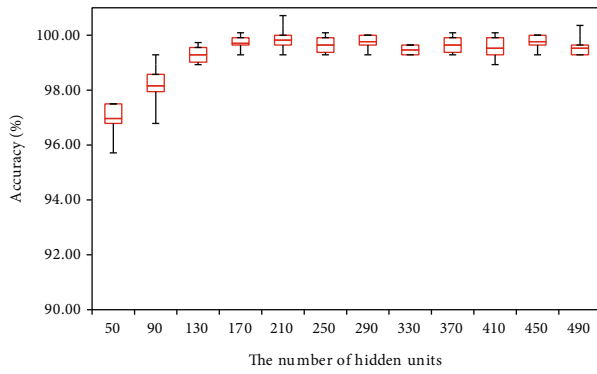
FIGURE 9: Boxplot of prediction accuracies with different numbers of hidden units.
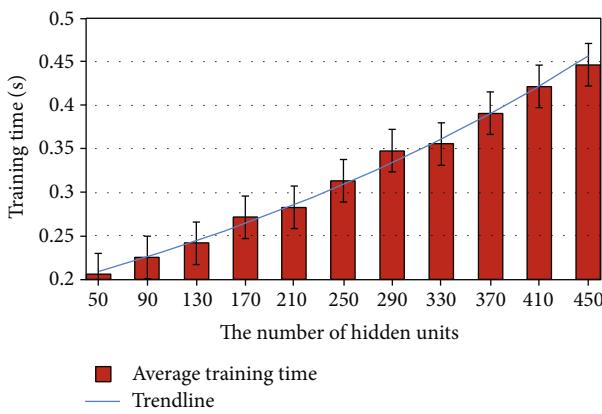

FIGURE 10: Training time of M_ELMConvNet with different numbers of hidden units.

FIGURE 11: Confusion matrix of multistatus classification and recognition accuracy based on M_ELMConvNet.
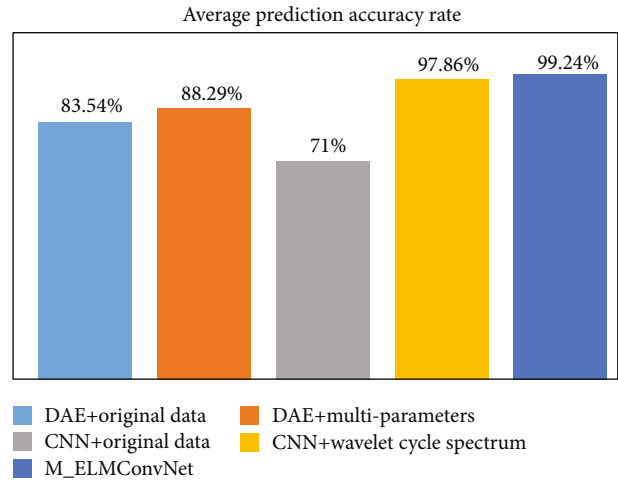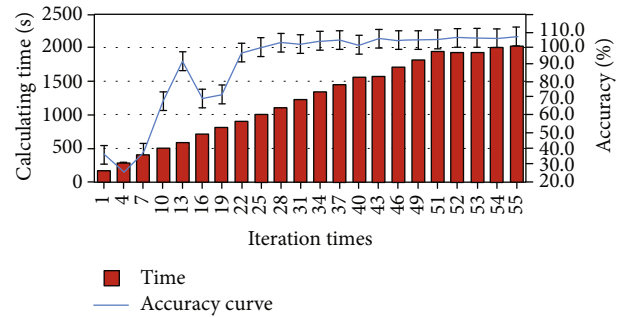

FIGURE 12: Evaluate classifiers' performance by comparison.

FIGURE 13: The result of different iteration times of the CNN model based on BP for planetary bearing status recognition.

distinction between status 2 and status 3. In general, the proposed method in this paper can achieve high predictive recognition accuracy in each status.

4.3. Comparative Verification. The diagnosis performances of the original data and multiparameters DAE and CNN with the original data and wavelet cycle spectrum were also compared with the M_ELMConvNet proposed in this paper. Each model was executed multiple times, and the average result was calculated. Figure 12 shows that the M_ELM-ConvNet achieved the highest average prediction accuracy of 99.24%. Thus, the proposed algorithm in this paper has strong noise suppression capability in the identification and classification of weak fault statuses.

In order to further verify the superiority of the proposed algorithm in computing time, the CNN model based on EBP was used as a comparison model to analyze the same sample data set with M_ELMConvNet. The results are shown in Figures 10 and 13. As can be seen from the figure, the calculation time of the algorithm increases linearly with the increase in the number of hidden units and the number of iterations. At the same time, the recognition accuracy of the algorithm gradually improves and tends to be stable. Considering the influence of the number of hidden units and the number of iterations on the recognition accuracy, the calculation time of 410 hidden units and 28 iterations was compared. One was 0.43 s, and the other was about 2000 s. The time difference is several orders of magnitude, which proved the superiority of the M_ELMConvNet algorithm in computational efficiency.

5. Conclusions and Further Works

In this paper, a new deep feature extraction and diagnosis method was proposed to improve the recognition accuracy and reduce computational complexity for weak failure signal of planetary bearing with large data volume. In M_ELM-ConvNet, ELM was embedded in the CNN model instead of convolution operation to avoid a repeated EBP operation process. After two processes of ELM feature dimensionality reduction and extraction, the amount of calculation was reduced and the prediction accuracy was improved. In addition, based on the parameter transfer theory, the model parameters extracted from the labeled training sample data are introduced into the unlabeled sample data training model to achieve prediction. The effectiveness and superiority are proven on experiment setup testing data. Moreover, analysis results show that the proposed model has advantages in recognition accuracy and operation speed compared with other methods.

The present work is mainly carried out in the case of sufficient sample data. However, in the actual operation of planetary bearings, the sample size of analysis data is unbalanced; that is, the trouble-free sample size is large, while the failure sample size is small. For failure data, such as in the early stage or in the case of large external noise, manual marking often leads to missed diagnosis or misdiagnosis. How to realize the self-supervised learning of unlabeled data and make it able to automatically extract data features and perform labeling is the work to be done in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] C. T. Yiakopoulos, K. C. Gryllias, and I. A. Antoniadis, "Rolling element bearing fault detection in industrial environments based on a K-means clustering approach," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2888–2911, 2011.

[2] J. Xu, J. Zhang, T. Chen, and R. H. Liu, "A planet bearing fault case of wind turbine gearbox," *Advanced Materials Research*, vol. 953-954, pp. 472–477, 2014.

[3] D. Zhao, S. Liu, D. Gu et al., "Enhanced data-driven fault diagnosis for machines with small and unbalanced data based on variational auto-encoder," *Measurement Science and Technology*, vol. 31, no. 3, article 035004, 2020.

[4] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 72-73, pp. 303–315, 2016.

[5] Q. Wang, C. Yang, H. Wan, D. Deng, and A. K. Nandi, "Bearing fault diagnosis based on optimized variational mode decomposition and 1D convolutional neural networks," *Measurement Science and Technology*, vol. 32, no. 10, article 104007, 2021.

[6] Y. Gao, X. Liu, and J. Xiang, "FEM simulation-based generative adversarial networks to detect bearing faults," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4961–4971, 2020.

[7] X. Y. Liu, H. Z. Huang, and J. W. Xiang, "A personalized diagnosis method to detect faults in gears using numerical simulation and extreme learning machine," *Knowledge-Based Systems*, vol. 195, article 105653, 2020.

[8] Z. H. An, X. X. Jiang, J. Cao, R. Yang, and X. G. Li, "Self-learning transferable neural network for intelligent fault diagnosis of rotating machinery with unlabeled and imbalanced data," *Knowledge-Based Systems*, vol. 230, article 107374, 2021.

[9] M. M.M. Islam and J. Kim, "Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network," *Computers in Industry*, vol. 106, pp. 142–153, 2019.

[10] X. L. Pei, X. Zheng, and J. L. Wu, "Intelligent bearing fault diagnosis based on Teager energy operator demodulation and multiscale compressed sensing deep autoencoder," *Measurement*, vol. 179, article 109452, 2021.

[11] D. D. Cheng, L. J. Liu, and Z. Yu, "CNN-based intelligent fault-tolerant control design for turbofan engines with actuator faults," *IEEE Access*, vol. 9, pp. 28122–28139, 2021.

[12] B. Zhao, X. Zhang, Z. Zhan, and Q. Wu, "A robust construction of normalized CNN for online intelligent condition monitoring of rolling bearings considering variable working conditions and sources," *Measurement*, vol. 174, article 108973, 2021.

[13] L. Fu, L. Zhang, and J. Tao, "An improved deep convolutional neural network with multiscale convolution kernels for fault diagnosis of rolling bearing," *Materials Science and Engineering*, vol. 1043, no. 5, article 052021, 2021.

[14] J. Zhao, S. Yang, Q. Li, Y. Liu, X. Gu, and W. Liu, "A new bearing fault diagnosis method based on signal-to-image mapping and convolutional neural network," *Measurement*, vol. 176, article 109088, 2021.

[15] Z. Chen, A. Mauricio, W. Li, and K. Gryllias, "A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks," *Mechanical Systems and Signal Processing*, vol. 140, article 106683, 2020.

[16] Z. Ye and J. Yu, "Deep morphological convolutional network for feature learning of vibration signals and its applications to gearbox fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 161, article 107984, 2021.

[17] G. Huang, G. B. Huang, S. Song, and K. You, "Trends in extreme learning machines: a review," *Neural Networks*, vol. 61, pp. 32–48, 2015.

[18] L. Kasun, H. Zhou, G. B. Huang, and M. V. Chi, "Representational learning with ELMs for big data," *IEEE Intelligent Systems*, vol. 6, pp. 31–34, 2013.

[19] J. Tang, C. Deng, and G. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Transactions on Neural Network and Learning System*, vol. 27, no. 4, pp. 809–821, 2016.

[20] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 7, pp. 1693–1702, 2017.

[21] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network,"

Measurement Science and Technology, vol. 26, no. 11, article 115002, 2015.

[22] C. Che, H. Wang, Q. Fu, and X. Ni, "Intelligent fault prediction of rolling bearing based on gate recurrent unit and hybrid autoencoder," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 235, no. 6, pp. 1106–1114, 2021.

[23] X. Fang, Y. Cai, Z. Cai, X. Jiang, and Z. Chen, "Sparse feature learning of hyperspectral imagery via multiobjective-based extreme learning machine," *Sensors*, vol. 20, no. 5, p. 1262, 2020.

[24] Y. Cai, X. Liu, W. Yu, H. Peng, and Z. Cai, "Extreme learning machine based on evolutionary multi-objective optimization," in *International Conference on Bio-Inspired Computing: Theories and Applications*, pp. 420–435, Springer Singapore, 2017.

[25] Y. Wu, Y. Zhang, X. Liu, Z. Cai, and Y. Cai, "A multiobjective optimization-based sparse extreme learning machine algorithm," *Neurocomputing*, vol. 317, pp. 88–100, 2018.

[26] L. L. C. Kasun, Y. Yang, G. B. Huang, and Z. Zhang, "Dimension reduction with extreme learning machine," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3906–3918, 2016.

[27] X. B. Wang, X. Zhang, Z. Li, and J. Wu, "Ensemble extreme learning machines for compound-fault diagnosis of rotating machinery," *Knowledge-Based Systems*, vol. 188, article 105012, 2020.

[28] S. Pang, X. Yang, X. Zhang, and Y. Sun, "Fault diagnosis of rotating machinery components with deep ELM ensemble induced by real-valued output-based diversity metric," *Mechanical Systems and Signal Processing*, vol. 159, article 107821, 2021.

[29] G. Du, Q. Xu, and X. Yang, "Fault diagnosis of rotating machinery components using a deep kernel extreme learning machine under different working conditions," *Measurement Science and Technology*, vol. 31, no. 11, article 115901, 2020.

[30] K. Li, M. Xiong, F. Li, L. Su, and J. Wu, "A novel fault diagnosis algorithm for rotating machinery based on a sparsity and neighborhood preserving deep extreme learning machine," *Neurocomputing*, vol. 350, pp. 261–270, 2019.

[31] J. Ou, H. Li, G. Huang, and G. Yang, "Intelligent analysis of tool wear state using stacked denoising autoencoder with online sequential-extreme learning machine," *Measurement*, vol. 167, article 108153, 2021.

[32] Z. Chen, K. Gryllias, and W. Li, "Mechanical fault diagnosis using convolutional neural networks and extreme learning machine," *Mechanical Systems and Signal Processing*, vol. 133, article 106272, 2019.

[33] R. Yang, H. Li, C. Wang, and C. He, "Rolling element bearing weak feature extraction based on improved optimal frequency band determination," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 203-210, pp. 1989–1996, 2018.

[34] L. Zhang, J. Li, and H. Lu, "Saliency detection via extreme learning machine," *Neurocomputing*, vol. 218, pp. 103–112, 2016.