Neurocomputing 433 (2021) 237–251

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Multi-stage fault diagnosis framework for rolling bearing based on OHF Elman AdaBoost-Bagging algorithm

Tangbin Xia [a], Pengcheng Zhuo [a], Lei Xiao [b], Shichang Du [a], Dong Wang [a], Lifeng Xi [a,*]

[a] State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, SJTU-Fraunhofer Center, Shanghai 200240, China
[b] College of Mechanical Engineering, Donghua University, Shanghai 201620, China

ARTICLE INFO

Article history:
Received 1 March 2020
Revised 8 September 2020
Accepted 2 October 2020
Available online 8 December 2020

Keywords:
Rolling bearing
OHF Elman AdaBoost-Bagging
Neural network
Multi-stage fault diagnosis

ABSTRACT

With the increasing complexity of industrial equipment, it is urgent to provide timely diagnosis and accurate evaluation to avoid failure. For rolling bearings, it is important to achieve the multi-stage (incipient, intermediate, late) fault diagnosis under random noise. Different from traditional methods, an Output Hidden Feedback Elman Adaptive Boosting-Bootstrap Aggregating algorithm is proposed under a comprehensive diagnosis framework. First, the original signal is decomposed, denoised and reconstructed by Ensemble Empirical Mode Decomposition. Then, OHF Elman neural network is designed by increasing a feedback from the output layer to the hidden layer based on Elman neural network. This improves the memory function for dynamic data of rolling bearings. Furthermore, for achieving diagnostic accuracy and algorithm stability, OHF Elman AdaBoost-Bagging algorithm is developed as a strong learner through the dual integration of AdaBoost algorithm and Bagging algorithm. Experimental results show that the proposed algorithm not only has a good diagnostic performance on different stages of rolling bearing faults, but also achieves higher generalization ability and stability. This multi-stage fault diagnosis framework provides a novel tool and an effective solution for rolling bearing fault diagnosis.

1. Introduction

In modern industry, an effective diagnosis framework is important to ensure equipment operation and avoid machine failure. However, due to the increasing structural complexity, an industrial equipment normally undergoes a long and multi-stage degradation, which brings new challenges for the fault diagnosis. In this situation, the different failure information of various multi-stage (incipient, intermediate, late) processes should be comprehensively utilized to provide timely diagnosis and accurate evaluation [1]. As an important kind of components, rolling bearings and their health evaluations are essential to ensure the equipment operation. Performance degradation or failure of rolling bearings will lead to unplanned equipment shutdown, resulting in economic loss and even heavy casualties [2]. Hence, it is pivotal and urgent to develop a multi-stage diagnosis framework to provide early warning and precise diagnosis for rolling bearing fault.

For developing such a multi-stage diagnosis framework, it should be noticed that rolling bearing failures are often reflected in the vibration signals. However, in the incipient stage of rolling bearing failures, such as early fatigue peeling, it is very difficult to conduct the prognostic and health management (PHM) because of the vague fault information and inconspicuous abnormal vibration signals [3]. If the early failure cannot be detected timely, it may result in a disastrous accident at a certain time point [4]. Therefore, effective condition monitoring and fault diagnosis techniques for rolling bearings in the incipient, intermediate and late stages is practicable and required, which is of great significance for improving manufacturing stability and equipment health management [5–8]. Thus, performing fault diagnosis for rolling bearings becomes more interesting because it is likely to benefit from utilizing diverse health information of incipient, intermediate, late stages to provide timely diagnosis and accurate evaluation.

In the diagnosis framework, the first issue is to handle the fault information monitored from the rolling bearing, which is the base for the following multi-stage fault diagnosis. It can be noted that the vibration signal of a rolling bearing is usually contaminated by noise and other vibration sources and consequently, the fault information may be thus masked. For handling this situation, commonly used denoising methods include Hilbert-Huang transform (HHT) [9], Wavelet transform (WT) [10], Empirical Mode Decomposition (EMD) [11] and Ensemble Empirical Mode Decomposition (EEMD) [12], etc. Among them, EEMD has a good denoising effect

* Corresponding author.
E-mail address: lfxi@sjtu.edu.cn (L. Xi).

https://doi.org/10.1016/j.neucom.2020.10.003
0925-2312/

Nomenclature

PHM	Prognostic and health management
HHT	Hilbert-Huang transform
WT	Wavelet transform
EMD	Empirical Mode Decomposition
EEMD	Ensemble Empirical Mode Decomposition
SVM	Support Vector Machine
ANN	Artificial Neural Network
RNN	Recurrent Neural Network
CNN	Convolutional Neural Network
OHF	Output Hidden Feedback
Bagging	Bootstrap Aggregating
IMF	Intrinsic Mode Function
EDM	Electro-discharge machining
MSE	Mean Square Error
$\boldsymbol{u}$	Input of neural network
r	Dimension of$\boldsymbol{u}$
$\boldsymbol{x}$	Output of the hidden layer
$\boldsymbol{x}_c$	Output of the context layer
n	Dimension of $\boldsymbol{x}_c$ and$\boldsymbol{x}$
$\boldsymbol{y}$	Output of the output layer
$\boldsymbol{y}_c$	Output of the context layer 2
m	Dimension of $\boldsymbol{y}$ and $\boldsymbol{y}_c$
$\boldsymbol{w}^{l1}$	Weight matrix of $\boldsymbol{x}_c$
$\boldsymbol{w}^{l2}$	Weight matrix of $\boldsymbol{u}$
$\boldsymbol{w}^{l3}$	Weight matrix of $\boldsymbol{x}$
$\boldsymbol{w}^{l4}$	Weight matrix of $\boldsymbol{y}_c$
$f(t)$	Activation function of the hidden layer
$g(v)$	activation function of the output layer
$\boldsymbol{y}_d$	desired output of neural network
$E(k)$	Error function in k-th iteration
η_1	learning step of $\boldsymbol{w}^{l1}$
η_2	learning step of $\boldsymbol{w}^{l2}$
η_3	learning step of $\boldsymbol{w}^{l3}$
η_4	learning step of $\boldsymbol{w}^{l4}$
f'	Derivative function of $f(t)$
g'	Derivative function of $g(v)$
α	self-feedback factor in the context layer
γ	self-feedback factor in the context layer 2
T	The original sample set
X_i	Input of T
Y_i	Output of T
N	Dimension of T
$H(x)$	Weak learner
K	Number of weak learners in Bagging
l	number of weak learners in AdaBoost
$G(x)$	Strong learner
E_j	the maximum error of j-th weak learner
$\widehat{E}_{jt}$	Relative error on t-th sample of j-th weak learner
O_{jt}	Output of j-th weak learner on t-th sample
e_j	Error rate of j-th weak learner
α_j	Weight coefficient of j-th weak learner
$w_j(t)$	Weight of t-th sample in j-th round
error_t	absolute error of t-th sample
τ	Threshold
b007	Fault with diameters of 0.007 in. in ball
b014	Fault with diameters of 0.014 in. in ball
b021	Fault with diameters of 0.021 in. in ball
IR007	Fault with diameters of 0.007 in. in inner race
IR014	Fault with diameters of 0.014 in. in inner race
IR021	Fault with diameters of 0.021 in. in inner race
OR007	Fault with diameters of 0.007 in. in outer race
OR014	Fault with diameters of 0.014 in. in outer race
OR021	Fault with diameters of 0.021 in. in outer race
$x(t)$	The raw vibration signal
$C_j(t)$	the j-th IMF component
N	The amount of white noise
r_n	The n-th residual component
N_{std}	Ratio of the standard deviation of the added white noise to the standard deviation of $x(t)$
$s(t)$	White noise
$\widehat{x}(t)$	Signal after adding $s(t)$
s	dimension of each sample
$\overline{MSE}$	Average error of the test data
STD	Standard deviation of MSE on the test data
a	An integer constant from 1 to 10

on vibration signals of rolling bearings, and thus can effectively extract incipient fault information. By adding Gaussian white noise into the original signal, the signals at different scales will have continuity. It helps to promote anti-aliasing decomposition, and effectively avoid mode aliasing and endpoint effect in EMD decomposition [12].

The key issue in the framework is the multi-stage fault diagnosis algorithm. Statistical models are the core of rolling bearing fault diagnosis, which can be divided into model-based approaches and data-driven approaches. Model-based approaches, as traditional approaches, are not applicable for large systems because of their complex physical models. Data-driven approaches have gained increasing attentions with the development of sensor technologies and data analysis methods [13]. At present, there are few studies on the diagnosis of rolling bearings in multiple fault stages, and some researches have applied vibration signal analysis and intelligent diagnosis methods in incipient fault diagnosis of rolling bearings. Commonly used methods for incipient fault diagnosis include probability statistics [14–17], Support Vector Machine (SVM) [18] and Artificial Neural Network (ANN). The representative work [19] solved the problem of early fault detection through deep neural network (DNN) and long short-term memory (LSTM) network, and proved the effectiveness and the reliability of the model on the bearing data sets. Li et al. [20] successfully recognized the early fault types of rolling bearings through the binary tree support vector machine (BT-SVM). Naïve Bayes classifier was designed to identify the most prominent fault for multiple incipient faults in transmission lines [21].

Due to their self-adaptive ability, massive parallel computations ability and universal estimations in both supervised and unsupervised ways, ANNs are widely used for PHM in industry [22]. Nowadays, ANNs have been developed vigorously, thus giving rise to many optimized neural networks, such as Elman Neural Network [23], Recurrent Neural Network (RNN) [24] and Convolutional Neural Network (CNN) [25]. Shi et al. [26] proposed OHF Elman (Output Hidden Feedback Elman) neural network based on Elman neural network by increasing a feedback between the output layer and the hidden layer. Most existing approaches employ individual learners designed for solving binary-class problems [27]. For the multiple fault diagnosis of rolling bearings, such a classification system is inefficient, and will exhibit low efficiency and large error for handling multi-classification problems [28,29]. Therefore, it is required to develop an effective technique for the multi-stage fault diagnosis of rolling bearings, which should consider each classifier's accuracy, diversity and independence.

Recently, numerous studies have been conducted to enhance the performance of the learners. Among them, Bootstrap Aggregating (Bagging) and Boosting [30] in ensemble learning are utilized

to design better learning approaches [31,32]. Lin et al. [33] integrated Bagging ensemble learning to optimize the SVM prediction model. Li et al. [34] evaluated the levels of the internal leakage of the hydraulic cylinders by using AdaBoost-BP neural network classifier. Bagging algorithm integrates different learners that adapt to different samples by changing the training set, thereby improving the diagnostic accuracy of the full sample. AdaBoost algorithm integrates different learners and corresponding weights by changing the sample weights, thereby forming a strong learner. Since the influence of each sample on diagnostic error is considered, AdaBoost algorithm has a better effect than Bagging algorithm on improving the diagnostic accuracy. However, AdaBoost is particularly sensitive to anomalous samples. Bagging algorithm can effectively reduce the impact of anomalous samples on AdaBoost algorithm by changing the training set. Due to the similarity of different fault characteristics of rolling bearings, a single learner cannot achieve effective classification of different fault samples. Hence, in order to ensure the applicability of the learner to the full sample data, this paper proposes a novel ensemble learning concept which integrates both Bagging and AdaBoost. A strong learner based on OHF Elman neural network (a weak learner) and AdaBoost-Bagging integrated ensemble learning is designed. Major fault components (rolling element, inner race and outer race) and their different fault stages (incipient, intermediate and late stages) are comprehensively analyzed to prove the effectiveness of this proposed method.

The remainder of this paper is organized as follows: Section 2 illustrates the multi-stage fault diagnosis of rolling bearings. Section 3 presents detailed mathematical formulations of the proposed technical framework and OHF Elman AdaBoost-Bagging Algorithm. Then, an experiment setup on a benchmark dataset is shown in Section 4. Section 5 provides the results comparison and effect analysis. Finally, Section 6 gives some concluding remarks and future work plans.

2. Problem description

In order to effectively grasp the fault status of rolling bearings, the multi-stage fault diagnosis is proposed and performed through vibration signal analysis. Rolling bearing fault diagnosis needs not only to accurately identify the fault location, but also to diagnose the fault stage. This study comprehensively analyzes three major fault components: rolling element (i.e. ball), inner race, outer race and three fault stages: incipient, intermediate, late stages. Hence, the multi-stage fault diagnosis of rolling bearings can be transformed into a multi-classification problem. The classification diagram is shown in Fig. 1.

The incipient fault diagnosis of rolling bearings has gained more and more attention, due to the weakness of the vibration signal with high background noise. Recently, some works have also made some progress in the incipient fault diagnosis of rolling bearings, for example, stochastic resonance method [1], EMD [4] and machine learning [20] have been well applied. However, two major problems remain to be settled:

1) In many cases, the impact signals related to bearing failure usually show different waveform characteristics. Even for the same bearing, the shock signals at different fault stages will also show very different waveforms. A more flexible and adaptive algorithm is of great significance for the multi-fault diagnosis of rolling bearings.
2) The performance of a diagnostic model is affected by the fault characteristics extracted from data and its own parameters. The same diagnostic algorithm model cannot guarantee the applicability of the data set under different faults and different working conditions.

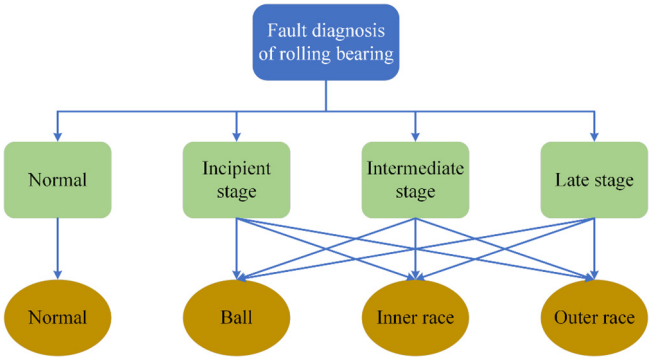


Fig. 1. Classification diagram of rolling bearing faults.

In order to retain the fault information of rolling bearings to the greatest extent, this study uses EEMD to perform the signal denoising. Taking the EEMD decomposition waveform of a normal signal consists of 2000 sample points as an example, the result is shown in Fig. 2.

From top to bottom, the 10 waveforms are respectively 9 Intrinsic Mode Function (IMF) components and a residual component R decomposed from 2000 sample points. It can be seen from Fig. 2 that the modal aliasing is substantially eliminated between the IMF components, and the residual component R is monotonically decreasing.

3. Methodology

In order to build a more flexible, adaptive and higher applicable algorithm, OHF Elman neural network is selected as an effective method for fault diagnosis, while a strong learner is developed by integrating AdaBoost and Bagging algorithms to perform more accurate fault diagnosis on rolling bearings. In addition, EEMD was utilized for modal decomposition and signal reconstruction. On this basis, this constructed technical framework is shown in Fig. 3.

The framework illustrated in Fig. 3 is mainly composed of fault signal extraction and fault mode identification. In terms of fault signal extraction, the kurtosis and correlation coefficient will be the judgment criteria for the selection of IMFs obtained through EEMD. Regarding fault mode identification, a strong learner is developed by integrating AdaBoost algorithm with Bagging algorithm based on OHF Elman neural network. Multiple training subsets are obtained by bootstrap method. Based on each training subset, OHF Elman neural network is performed and OHF Elman AdaBoost learner is obtained. Finally, the OHF Elman AdaBoost learners from all training subsets are combined to be the OHF Elman AdaBoost-Bagging learner.

The description of this technical framework is presented here. The first part is the introduction of EEMD for Signal processing and feature extraction, after that, the network architecture of OHF Elman neural network will be illustrated. Then, the Bagging and AdaBoost algorithm will be performed in detail. Finally, the specific iterative process of OHF Elman AdaBoost-Bagging Algorithm is provided.

3.1. EEMd

The principle of EEMD decomposition is [35]: When the additional white noise is evenly distributed throughout the time-frequency space, the time-frequency space is composed of different

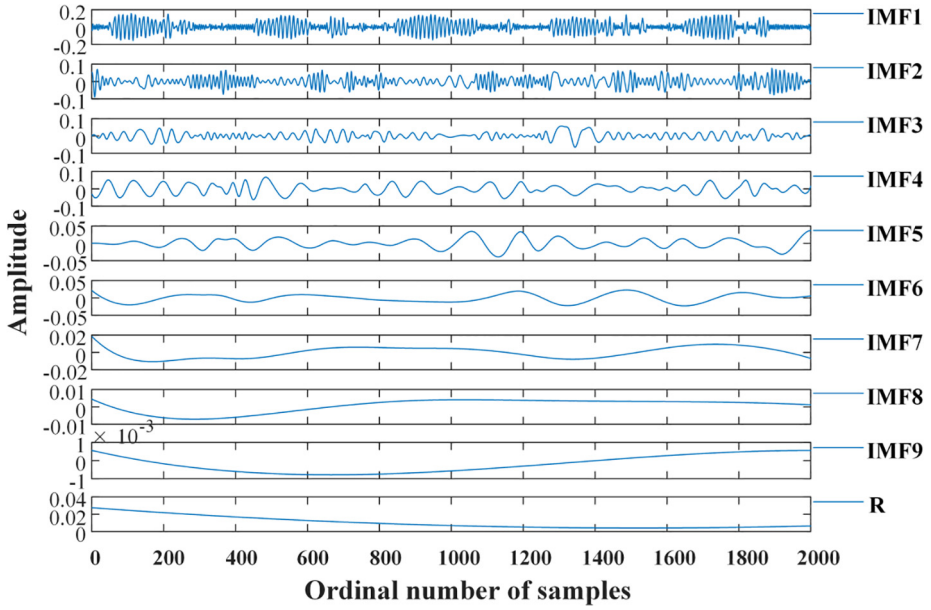


Fig. 2. Decomposition waveform of normal signal.

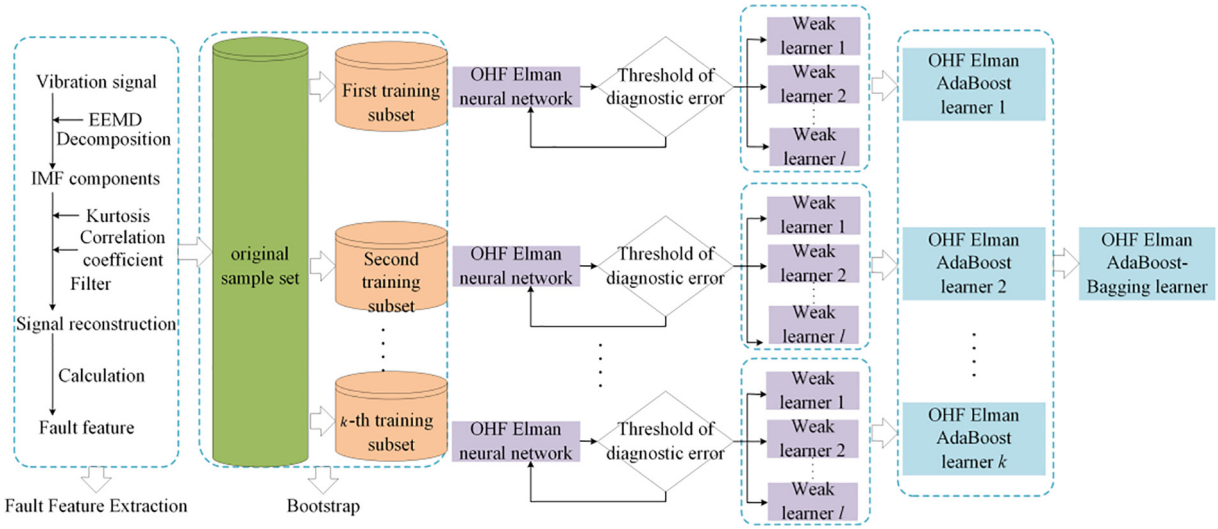


Fig. 3. Technical framework of fault diagnosis for rolling bearing.

scale components divided by the filter bank. EEMD has an additional step of adding noise on the basis of EMD. The white noise is evenly distributed on the time-frequency space of the signal. When the number of additions is sufficient, the noise effect of the signal can be eliminated, and the average of the entire IMF component can be obtained. This is the effective and accurate IMF component.

The EEMD decomposition process is as follows:

Step 1: Adding the white noise $s(t)$ to the original signal $x(t)$:

$$\hat{x}(t) = x(t) + s(t) \tag{1}$$

Step 2: EMD decomposition of $\hat{x}(t)$:

$$C_1(t) = \hat{x}(t) - r_1(t) \tag{2}$$

$$C_2(t) = r_1(t) - r_2(t) \tag{3}$$

$$\vdots$$

$$C_n(t) = r_{n-1}(t) - r_n(t) \tag{4}$$

where the n-th residual component $r_n(t)$ satisfies the characteristics of the monotone function, stopping the EMD decomposition, that is, $r_n(t)$ is the final residual component.

Step 3: Repeating Step 1 and Step 2 N' times.

Step 4: Averaging the N' groups of IMFs components from N' times EMD decompositions.

$$C_j(t) = \frac{1}{N'} \sum_{i=1}^{N'} C_{ij}(t) \tag{5}$$

$$x(t) = \sum_{j=1}^{n} C_j(t) + r_n \tag{6}$$

$x(t)$ represents the raw vibration signal; $C_j(t)$ represents the j-th IMF component; N' is the amount of white noise, usually 50 or 100; n is the number of IMF components; r_n is the residual component; N_{std} is the ratio of the standard deviation of the added white noise to the standard deviation of the raw vibration signal.

3.2. Weak Learner: Neural network

OHF Elman achieves the ability to process dynamic data by adding feedback from the output layer to the hidden layer based on Elman neural network. In order to illustrate the effectiveness of the feedback structure, this study will first conduct a comparison experiment between the Elman neural network and the OHF Elman neural network. And their specific mathematical models are described here.

3.2.1. Elman neural network

Assume that there are r nodes in the input layer, n nodes in the hidden and context layers, respectively, and m nodes in the output layer. Then the input $\boldsymbol{u}$ is an r-dimensional vector, the output $\boldsymbol{x}$ of the hidden layer and the output $\boldsymbol{x}_c$ of the context nodes are n-dimensional vectors, respectively, while the output $\boldsymbol{y}$ of the output layer is an m-dimensional vector. The weights $\boldsymbol{w}^{l1}$, $\boldsymbol{w}^{l2}$, and $\boldsymbol{w}^{l3}$ are $n \times n$, $n \times r$ and $m \times n$-dimensional matrices, respectively. These weights of the context nodes, input nodes and hidden nodes are denoted as $\boldsymbol{w}^{l1}$, $\boldsymbol{w}^{l2}$, and $\boldsymbol{w}^{l3}$.

The modified Elman network adds a self-feedback factor α in the context nodes, based on the traditional Elman neural network [26]. Its mathematical model is:

$$\boldsymbol{x}(k) = f\left(\boldsymbol{w}^{l1}\boldsymbol{x}_c(k) + \boldsymbol{w}^{l2}\boldsymbol{u}(k-1)\right) \tag{7}$$

$$\boldsymbol{x}_c(k) = \alpha\boldsymbol{x}_c(k-1) + \boldsymbol{x}(k-1) \tag{8}$$

$$\boldsymbol{y}(k) = g\left(\boldsymbol{w}^{l3}\boldsymbol{x}(k)\right) \tag{9}$$

Here, $f(t)$ is often taken as the sigmoid function:

$$f(t) = \frac{1}{1+e^{-t}} \tag{10}$$

while $g(v)$ represents the activation function of the output layer and is often taken as a linear function. If $g(v)$ is taken as a linear function, then $g_i' = 1$. And α is in a range of (0, 1).

Let the kth desired output of the system be $\boldsymbol{y}_d(k)$. The error function will be:

$$E(k) = \frac{1}{2}(\boldsymbol{y}_d(k) - \boldsymbol{y}(k))^T(\boldsymbol{y}_d(k) - \boldsymbol{y}(k)) \tag{11}$$

Differentiating $E(k)$ with respect to $\boldsymbol{w}^{l3}$, $\boldsymbol{w}^{l2}$, and $\boldsymbol{w}^{l1}$, respectively, according to the gradient descent method, we get the following equations:

$$\Delta w_{ij}^{l3} = \eta_3\delta_i^0 x_j(k), i = 1,2,\cdots,m; j = 1,2,\cdots,n \tag{12}$$

$$\Delta w_{jq}^{l2} = \eta_2\delta_j^h u_q(k-1), j = 1,2,\cdots,n; q = 1,2,\cdots,r \tag{13}$$

$$\Delta w_{jl}^{l1} = \eta_1\sum_{i=1}^{m}\left(\delta_i^0 w_{ij}^{l3}\right)\frac{\partial x_j(k)}{\partial w_{jl}^{l1}}, j = 1,2,\cdots,n; l = 1,2,\cdots,n \tag{14}$$

where η_3, η_2,η_1 are learning steps of $\boldsymbol{w}^{l3}$, $\boldsymbol{w}^{l2}$, and $\boldsymbol{w}^{l1}$, respectively, and

$$\delta_i^0 = (y_{d,i}(k) - y_i(k))g_i' \tag{15}$$

$$\delta_j^h = \sum_{i=1}^{m}\left(\delta_i^0 w_{ij}^{l3}\right)f_j' \tag{16}$$

$$\frac{\partial x_j(k)}{\partial w_{jl}^{l1}} = x_l(k-1)f_j' + \alpha\frac{\partial x_j(k-1)}{\partial w_{jl}^{l1}} \tag{17}$$

3.2.2. OHF Elman neural network

OHF Elman neural network increases a feedback $\boldsymbol{y}_c(k)$ from the output layer$\boldsymbol{y}(k)$ to the hidden layer $\boldsymbol{x}_c(k)$, which is illustrated in Fig. 4.

The mathematical model of OHF Elman neural network is represented as:

$$x(k) = f\left(\boldsymbol{w}^{l1}\boldsymbol{x}_c(k) + \boldsymbol{w}^{l2}\boldsymbol{u}(k-1)\right) \tag{18}$$

$$\boldsymbol{x}_c(k) = \alpha\boldsymbol{x}_c(k-1) + x(k-1) \tag{19}$$

$$\boldsymbol{y}_c(k) = \gamma\boldsymbol{y}_c(k-1) + \boldsymbol{y}(k-1) \tag{20}$$

$$\boldsymbol{y}(k) = g\left(\boldsymbol{w}^{l3}x(k) + \boldsymbol{w}^{l4}\boldsymbol{y}_c(k)\right) \tag{21}$$

whereγ $(0 < \gamma < 1)$ is the gain factor of the self-feedback. The connection weights of the context layer 2 nodes can be denoted as $\boldsymbol{w}^{l4}$. And$\boldsymbol{y}_c$ is the output of the context layer 2. The modifications on the weights$\boldsymbol{w}^{l1}$, $\boldsymbol{w}^{l2}$and$\boldsymbol{w}^{l3}$are identical to those in the modified Elman network, while the update rule on $\boldsymbol{w}^{l4}$ can be:

$$\Delta w_{ij}^{l4} = \eta_4\delta_i^0 y_{Cj}(k), i = 1,2,\cdots,m; j = 1,2,\cdots,\ n \tag{22}$$

where η_4 is the learning step of $\boldsymbol{w}^{l4}$and δ_i^0 is given by Eq. (15), while the context layers 2 nodes are equal to the output nodes.

3.3. Bagging algorithm

One integrated approach in our study is the Bagging algorithm, which is a typical representative of parallel ensemble learning method and is directly based on bootstrap method [36]. This bootstrap method can not only reduce the training complexity, but also guarantee the diversity of multiple weak learners. Bagging firstly uses training datasets to obtain multiple training subsets by bootstrap method, then uses multiple weak learners to learn each subsample set. Finally, the average of the multiple weak learners is the strong learner based on Bagging algorithm [37]. The core idea is to obtain diversity by building several weak learning models with multiple training subsets. We integrate this method since it is especially suitable for those unstable weak learners that are sensitive to few variations in the training set.

The flow chart is illustrated in Table 1 and the specific iterative process of the Bagging algorithm is as follows: (1) Carrying out K

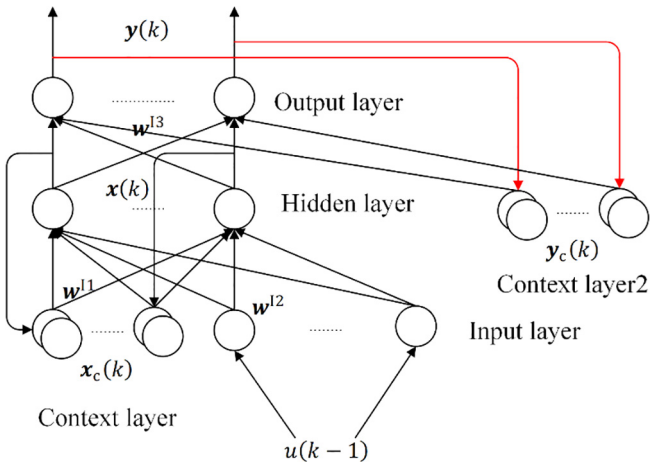


Fig. 4. Architecture of OHF Elman network.

Table 1
Flow chart of Bagging algorithm.

Input:	The original sample set: T:$\{(X_i, Y_i)\}_{i=1}^{N}$ Weak learner:$H(x)$ The number of weak learners: K
Process	
01:	For $i = 1, 2, \cdots, K$
02:	Extracting n training samples from sample set: T randomly as i-th training subset
03:	Using i-th training subset for $H(x)$, $H_i(x)$ will be obtained
04:	End for
05:	Combining weak learners using average method: $G(x) = \sum_{i=1}^{K} H_i(x)/K$
Output:	Strong learner: $G(x)$

rounds of the bootstrap method for the original sample set. In each round, extracting N_0 training samples. Finally, K groups of independent training subsets are obtained. (2) Learning the K groups of training subsets by using given weak learner respectively, then Kindependent weak learners will be developed. (3) Calculating the average of the output of the Kweak learners as the final result, i.e., the output of the strong learner.

3.4. AdaBoost algorithm

Another integrated approach here is AdaBoost, which is a machine learning integration algorithm to improve data training performance [38]. Generally, AdaBoost chooses a weak learner as the basic learner, such as classification trees, support vector machines and neural networks [39]. It emphasizes the sample with the incorrect learning effect or the large learning error by appropriately adjusting the weight in the training data distribution before the next weak learner training. Therefore, we utilize the AdaBoost algorithm, since the predictive diagnosis of test data can be improved by integrating each step of the weak learner.

Many methods have been used for modifying AdaBoost to solve regression problems. Currently, AdaBoost R2 [40,41] and AdaBoost RT [42] are most commonly used and effective. The AdaBoost algorithm is more flexible when building a strong learner and the results are less prone to overfitting. The flow chart of AdaBoost R2 algorithm is shown in Table 2.

Considering the diversity of bearing failures in reality, the category of bearing fault is expressed as $[0, \cdots, 0, 1, 0 \cdots, 0]^T$ (the position i corresponding to 1 represents the i-th fault mode), which can reflect the probability that the sample is divided into each type. The maximum error of AdaBoost R2 algorithm is no longer applicable, and a novel AdaBoost regression algorithm is designed.

The characteristics of the proposed AdaBoost algorithm:

1) Using a novel threshold to effectively split the sample into samples with larger errors and samples with smaller errors. When the threshold is determined reasonably, the model iteration will be accelerated by considering the samples with larger errors exclusively. The threshold here can be obtained from the diagnosis result of OHF Elman neural network.
2) Considering that the Bagging algorithm reduces the effect of singular samples on the iterative change of sample weights in the AdaBoost algorithm, the algorithm can use different constant coefficients to update the sample weights, thereby speeding up the iteration of sample weights Table 3.

3.5. The proposed OHF Elman AdaBoost-Bagging algorithm

As mentioned above, due to the non-stationarity and nonlinearity of vibration signals, and doped with different degrees of noise, the neural network is improved to perform the multi-stage

Table 2
Flow chart of AdaBoost R2 algorithm.

Input:	The original sample set: T:$\{(X_i, Y_i)\}_{i=1}^{N}$ Weak learner: $H(x)$ The number of weak learners: l
Process	
01:	Normalize weights: $w_1(t) = 1/N, t = 1, \cdots, N$
02:	For$j = 1, 2, \cdots, l$
03	Using the training set for $H(x)$, $H_j(x)$ will be obtained
04:	Calculating the maximum error: $E_j = \max_t \lvert O_{jt} - Y_{jt} \rvert, t = 1, \cdots, N$ O_{jt}: Output of$H_j(x)$ Calculating the relative error of each sample: Linear: $\hat{E}_{jt} = \frac{\lvert O_{jt} - Y_{jt} \rvert}{E_j}, t = 1, \ldots, N$ Squared Law: $\hat{E}_{jt} = \frac{(O_{jt} - Y_{jt})^2}{E_j^2}, \quad t = 1, \ldots, N$ Exponential: $\hat{E}_{jt} = 1 - e^{\frac{-\lvert O_{jt} - Y_{jt} \rvert}{E_j}}, \quad t = 1, \ldots, N$
05	Calculating the error rate: $e_j = \sum_{t=1}^{N} w_j(t) \times \hat{E}_{jt}$
06	Calculating the weight coefficient: $\alpha_j = \frac{e_j}{1 - e_j}$
07	Updating the weights: $w_{j+1}(t) = w_j(t) \times \alpha_j^{1 - \hat{E}_{jt}}, \quad t = 1, \ldots, N$ $w_{j+1}(t) = \frac{w_{j+1}(t)}{\sum_{t=1}^{N} w_{j+1}(t)}$
08:	End for
09:	Combining weak learners: $G(x) = G_{l^*}(x)$ $G_{l^*}(x)$: The weak learner corresponding to the sequence number l^* corresponding to the median value of $\ln \frac{1}{\alpha_j} j = 1, 2, \cdots, l$
Output:	Strong learner: $G(x)$

Table 3
Flow chart of a novel AdaBoost regression algorithm.

Input:	The original sample set: T:$\{(X_i, Y_i)\}_{i=1}^{N}$ Weak learner: $H(x)$ The number of base learners:l
Process	
01:	Normalize weights:$w_1(t) = 1/N, t = 1, \cdots, N$
02:	For$j = 1, 2, \cdots, l$
03	Using the training set for $H(x)$, $H_j(x)$ will be obtained
04:	Calculating the absolute error:$\text{error}_t = \sum_{h=1}^{10} \lvert O_{th} - Y_{th} \rvert, t = 1, \cdots, N$
05	Calculating the error rate:$e_j = \sum w_j(t), \text{error}_t > \tau$
06	Calculating the weight coefficient:$\alpha_j = \frac{1}{2} \ln \frac{1 - e_j}{e_j}$
07	Updating the weights: $w_{j+1}(t) = \begin{cases} w_j(t) \times 1.1, error_t > \tau \\ w_j(t), error_t \leqslant \end{cases}, t = 1, \ldots, N$ $w_{j+1}(t) = \frac{w_{j+1}(t)}{\sum_{t=1}^{N} w_{j+1}(t)}$
08	Normalizing the weight coefficient of weak learner:$\alpha_j = \frac{\alpha_j}{\sum_{j=1}^{K} \alpha_j}$
09:	End for
10:	Combining weak learners:$G(x) = \sum_{j=1}^{l} \alpha_j \times H_j(x)$
Output:	Strong learner:$G(x)$

fault diagnosis of rolling bearing, because the neural network is with good self-applicability and fault tolerance [43]. In order to further improve the diagnosis effect on the full sample data, ensemble learning is applied to the neural network. AdaBoost and Bagging ensemble algorithms have lower generalization error, and reduce the likelihood of overfitting. In general, the AdaBoost algorithm has higher accuracy than the Bagging algorithm [44]. However, we notice that the AdaBoost algorithm is sensitive to anomalous samples, because the anomalous samples may get larger weights in the iteration process, affecting the prediction accuracy of the final strong learner. Thus, we introduce Bagging ensemble to adjust the training set for AdaBoost ensemble. By adjusting the training set of the AdaBoost algorithm to subsets of the original training set, the effect of anomalous samples on the AdaBoost algorithm is reduced. Table 4 illustrates the specific iter-

Table 4
Flow chart of OHF Elman AdaBoost-Bagging Algorithm.

Input:	The original sample set: $T: \{(X_i, Y_i)\}_{i=1}^{N}$ Weak learner: $H(x)$-OHF Elman neural network The number of base learners for Bagging Ensemble: KThe number of base learners for AdaBoost Ensemble:l
Process	
01:	For $i = 1, 2, \cdots, K$
02:	Extracting n training samples from sample set: Trandomly as i-th training subsetT_i
03:	Normalizing weights:$w_1(t) = 1/N, t = 1, \cdots, N$
04:	For $j = 1, 2, \cdots, l$
05	Using T_i for $H(x)$, $H_{ij}(x)$ will be obtained
06:	Calculating the absolute error: $error_t = \sum_{h=1}^{10} \lvert O_{th} - Y_{th} \rvert$, $t = 1, \ldots, N$
07:	Calculating the error rate: $e_j = \sum w_j(t), \text{error}_t > \tau$
08:	Calculating the weight coefficient: $\alpha_j = \frac{1}{2} \ln \frac{1-e_j}{e_j}$
09:	Updating the weights: $w_{j+1}(t) = \begin{cases} w_j(t) \times 1.1, error_t > \tau \\ w_j(t), error_t \leqslant \end{cases}, t = 1, \ldots, N w_{j+1}(t) = \frac{w_{j+1}(t)}{\sum_{t=1}^{N} w_{j+1}(t)}$
10:	Normalizing the weight coefficient of weak learner:$\alpha_j = \frac{\alpha_j}{\sum_{j=1}^{l} \alpha_j}$
11	End for
12:	Strong learner based on AdaBoost:$G_i(x) = \sum_{j=1}^{l} \alpha_j \times H_j(x)$
13:	End for
14:	Combining learners:$G(x) = \sum_{i=1}^{K} G_i(x)/K$
Output:	Strong learner:$G(x)$

ation process of the proposed OHF Elman AdaBoost-Bagging algorithm.

The contributions of this OHF Elman AdaBoost-Bagging Algorithm can be concluded as:

1) *High-performance weak learner:* OHF Elman neural network is utilized by increasing a feedback from the output layer to the hidden layer. Hence, it has higher dynamic processing performance for bearing vibration signals. Besides, the superiority of OHF Elman neural network over SVM has been proven by the comparative data experiment.
2) *Multi-stage Fault Diagnosis of Rolling Bearing:* The fault data selected includes not only the data of the ball, the inner race and the outer race, but also their corresponding incipient, intermediate and late fault stages' health information. OHF Elman AdaBoost-Bagging Algorithm can achieve the diagnosis of the fault locations and stages of rolling bearings, thus more effective health management for rolling bearings can be performed.
3) *Higher generalization ability:* AdaBoost algorithm fully considers the diagnostic error of each sample data. And the weight coefficient of each weak learner will be determined by the training error. Through this path, the proposed method can effectively reduce the generalization error by integrating the trained weak learners.
4) *Higher stability model:* AdaBoost algorithm is extremely sensitive to abnormal data because it will later affect the generated weak learners. To handle this, Bagging algorithm is utilized to reduce the impact of abnormal data by bootstrap method. Meanwhile, Bagging algorithm can obtain accurate diagnosis models for different fault types by changing the training data. Therefore, the proposed OHF Elman AdaBoost-Bagging Algorithm has better stability when handling the data with high diversity.

4. Experimental study

The performances of this proposed OHF Elman AdaBoost-Bagging Algorithm is empirically evaluated on bearing vibration dataset from Case Western Reserve University Bearing Center. Firstly, descriptions of the dataset are presented. Then details of the experimental procedure are provided. At last, the proposed algorithms and evaluation criteria are designed.

4.1. Datasets description

The bearing vibration dataset from Case Western Reserve University Bearing Center has been widely used for rolling bearing fault diagnosis research [45]. This bearing dataset was obtained through experiments with 0.007, 0.014 and 0.021 in. diameter faults using SKF bearings. The motor speed is 1797RPM, and the sampling frequency is 12 kHz for the inter race, ball and outer race fault data (including normal mode). There are 10 fault modes in this study, selecting 120,000 sample points for each fault and 240,000 sample points for normal mode. The motor bearings were seeded with single-point faults using electro-discharge machining (EDM). 0.007, 0.014 and 0.021 in. are the fault diameters. Once a single-point fault occurs in the rolling bearing of the actual equipment, the fault diameter will increase with the development of the failure time. Therefore, this study chooses the faults with fault diameter of 0.007 in. as the incipient stage, 0.014 in. as the intermediate stage, and 0.021 in. as the late stage. In the following, b007, b014, and b021 represent rolling element (i.e. ball) faults with diameters of 0.007, 0.14, and 0.21 in.; IR007, IR014, and IR021 represent inner race faults; OR007, OR014, OR021 represent outer race faults.

4.2. Data processing

In the industry, bearing fault vibration signals usually appear as periodic transient pulses. In order to effectively display the characteristics in the time–frequency domain, the sampling signal normally covers at least two or three cycles. Considering the sampling frequency and characteristic frequency, we select 2000 sample points to represent the characteristics of the fault vibration signal. Every 2000 sample points are collected as a set of input data according to the time sequence. That is, there are 60 sets of sample data for each failure mode (covering 9 types of incipient, intermediate and late failures of the ball, inner race and outer race). Besides, there are 120 sets of sample data for the normal mode.

In this study, EEMD is used to achieve a good denoising effect on the vibration signals of rolling bearing. The IMF components are selected by using the kurtosis and correlation coefficient, thus the fault information is retained to the greatest extent. Then the fault characteristics of rolling bearing can be effectively extracted.

4.2.1. Signal reconstruction

The noise is completely independent of the original signal and the kurtosis value is more sensitive to the impact component of the signal [46]. Thus, the correlation coefficient and kurtosis value can be used as the indices of the IMF component selection [47]. In this paper, 0.1 is selected as the screening threshold of the correlation coefficient. IMF components with correlation coefficients higher than 0.1 are retained, while those lower than 0.1 are removed. IMF components reserved in different fault modes are not the same. IMF1–IMF5 are reserved for normal mode, while IMF1 and IMF2 are reserved for OR007 failure mode. By calculating the kurtosis values of the remaining IMF components, it is found that the kurtosis values of the IMF components retained from the inner race and outer race faults signal are high, and there is a large difference in various fault diameters. Then, selecting the first two IMF components with higher kurtosis values for signal reconstruction: normal mode selects IMF3 and IMF5; b007 and b021 fault mode select IMF1 and IMF3; b014 fault mode selects IMF1 and

IMF4; IR007, IR014, IR021, OR007 and OR021 fault mode select IMF1 and IMF2; OR014 fault mode selects IMF3 and IMF4.

4.2.2. Fault feature extraction

Selecting proper fault characteristics is the prerequisite for accurate fault diagnosis [18]. Through the analysis of EEMD reconstruction signal, the signal characteristic parameters can be calculated to determine whether the bearing has failed, and the type of fault. Selecting three dimensional parameters (mean, standard deviation, and root mean square value) and three dimensionless parameters (skewness, kurtosis, and margin), as shown in Table 5.

4.2.3. Input dataset and output dataset

In sum, the input of the sample data is composed of 6 kinds of time domain parameters extracted from the fault features. For each type of fault, 60 sets of sample data are selected, while 120 sets of sample data are selected for normal mode. The output of the sample data is represented by $[0,\cdots,0,1,0\cdots,0]^T$ (the position i corresponding to 1 represents the i-th fault mode), where normal, b007, b014, b021, IR007, IR014, IR021, OR007, OR014, OR021 represent the 1st to 10th fault modes, respectively. At the same time, the training data and test data are randomly split from the sample data according to a ratio of 5:1.

4.3. Performance metrics

Mean Square Error (MSE) is widely used as the index of performance evaluation for machine learning [48]. MSE is the expectation of the square of the difference between the estimated value of the parameter and the true value, which reflects the accuracy of diagnostic results:

$$MSE = \sum_{i=1}^{s}\left(y_{d,i} - y_i\right)^2/s \tag{23}$$

where $y_{d,i}$ is the real output, y_i is the output of model, and s is the dimension of each sample.

To characterize the stability of the model for rolling bearing fault diagnosis, the standard deviation is conducted for MSE of the test data. The smaller the standard deviation, the more balanced the diagnosis ability of the model for different faults [49].

$$STD = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(MSE_j - \overline{MSE}\right)^2} \tag{24}$$

where MSE_j is the error of the test data, while $\overline{MSE}$ is the average error of the test data and N is the number of the test data.

Although this study transforms the classification problem into a regression problem to adapt to the multi-fault characteristics of rolling bearings, MSE still cannot directly express the performance of a classification model. Therefore, the accuracy will remain as one of the performance metrics. According to the determined desired output:$[0,\cdots,0,1,0\cdots,0]^T$, each sample will result in an absolute error matrix, and the upper limit of each element in the absolute error matrix is defined as 0.5, that is, all 10 elements are less than 0.5 as correct diagnosis result. Both the signal processing and data analysis were conducted by MATLAB® (The MathWorks, Inc., Natick, MA, USA) (version: 9.5 (R2018b)).

5. Results and discussions

For achieving the fault diagnosis of rolling bearings in multiple fault stages and locations, this paper combines the idea of ensemble learning with OHF Elman neural network. To improve the generalization ability and stability of the method simultaneously, AdaBoost and Bagging algorithms are integrated to OHF Elman neural network.

5.1. Results analysis of different weak learners

As mentioned above, SVM and Neural Network are the most commonly used algorithms for fault diagnosis. In order to provide a more effective approach for rolling bearing fault diagnosis, we choose a better algorithm as the weak learner for ensemble learning. In this research, SVM and Neural Network are performed based on the same training data and test data.

On the one hand, since a small number of support vectors determine the final decision function of the SVM, the calculation depends on the support vector, rather than the entire sample space, and dimensional disaster can be avoided. This research is a regression problem for the fault diagnosis of rolling bearings. Therefore, the epsilon-SVR (support vector regression) model based on an insensitive loss function is used [50]. The purpose is to find a separating hypersurface, so that the expected error is minimized. The determination of the kernel function is crucial for the construction of epsilon-SVR model. In this study, RBF (Radial Basis Function) is selected because RBF can map samples to a higher-dimensional space. The determination of the gamma parameter of RBF and the penalty factor of epsilon-SVR still have a huge impact on the diagnosis results. Here, Cross-validation methods are usually used to improve the prediction accuracy. In this study, the penalty factor of epsilon-SVR is determined as 4 and the gamma parameter of RBF is determined as 0.25.

On the other hand, Elman neural network and OHF Elman neural network are both performed. In comparison with SVM, the superiority of OHF Elman compared with Elman neural network is also judged according to the data experiment. Tansig, logsig and trainlm are selected as the excitation function, input transfer function and network training function of neural network respectively. The number of neurons in the hidden layer can be between 4 and 14 by calculating:

$$n = \sqrt{r+m} + a \tag{25}$$

where$n$ is the number of neurons in the hidden layer, r is the number of neurons in the input layer, m is the number of neurons in the output layer, and a is an integer constant from 1 to 10.

Here, Elman neural network is performed under different numbers of neurons in the hidden layer. The number of neurons with the smallest generalization error (MSE) is selected for the hidden layer of the both two neural networks. As shown in Fig. 5, the number of neurons in the hidden layer is determined to be 10.

Furthermore, for selecting weak learner from SVM, we utilize Elman neural network and OHF Elman neural network as the basic model of ensemble learning. This study focuses on establishing the corresponding three models, and using the test dataset to analyze

Table 5
Three dimensional and three dimensionless parameters.

Dimensional parameters	Dimensionless parameters				
Mean	Standard deviation	Root mean square	Skewness	Kurtosis	Margin
$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$	$\sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$	$x_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$	$S = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^3}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2\right)^{3/2}}$	$K = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2\right)^2}$	$L = \frac{\max_i x_i - \min_i x_i}{\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert x_i\rvert}\right)^2}$

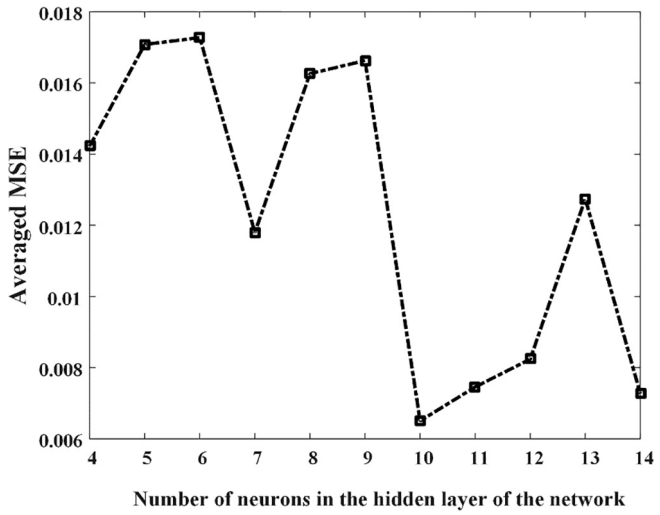


Fig. 5. Diagnostic error with different numbers of neurons in the hidden layer.

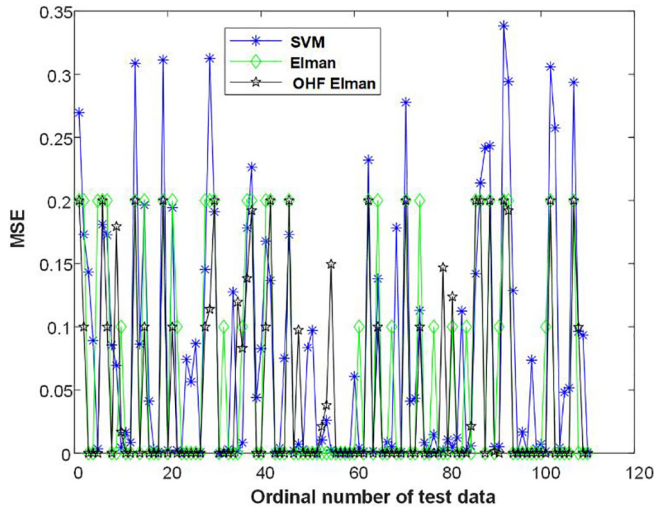


Fig. 6. Diagnostic errors of SVM and two neural networks on each sample.

the generalization abilities and stabilities. The MSE of each sample of the test dataset is shown in Fig. 6.

As can be seen from Fig. 6, Elman neural network and OHF Elman neural network have shown better diagnostic results on the test sample data, especially reducing the extremum of diagnostic error MSE. Compared with Elman neural network, OHF Elman neural network has no strong superiority. However, it has reduced diagnostic errors of many test samples. In order to better measure the pros and cons of these three models, the accuracy, the response time and the average and standard deviation of MSE of the test dataset are calculated, as shown in Table 6 and the Fig. 7.

First of all, the three algorithms have almost no difference in response time. When focusing on the entire test dataset, the accuracy, averaged MSE and standard deviation become the most effective analysis indicators. It can be noticed that Elman neural network and OHF Elman neural network have huge advantages on the entire sample, thus proving their better generalization ability and higher stability. Meanwhile, OHF Elman neural network has achieved the best results. Therefore, in terms of diagnostic results, OHF Elman neural network is the best choice for weak learner of ensemble learning.

Table 6
Diagnostic results of SVM, Elman and OHF Elman neural network.

Performance Metrics	SVM	Elman	OHF Elman
Accuracy	50%	64.5%	67.27%
Averaged MSE	0.0804	0.0609	0.0503
Standard Deviation	0.0978	0.0868	0.0768
Response Time/s	0.006	0.008	0.01

The key issue to establish an effective ensemble learning algorithm is to determine parameters of the model. In the previous section, the relevant parameters of the OHF Elman neural network have been obtained. Here, before performing specific results analysis, the relevant parameters of AdaBoost ensemble and Bagging ensemble need to be optimized.

5.2. Parameters determination of AdaBoost

The parameters of AdaBoost ensemble are the threshold τ and the number of base learners l. The parameter determination process and the results are as follows:

- Determination of the threshold: τ

The determination of the threshold τ directly determines the weight change of each sample, and thus affects the diagnostic error level of the strong learner. If this threshold is designed too small, the complexity and time of AdaBoost iterations will increase greatly. If it is too large, AdaBoost will lose its effect. To set a reasonable threshold τ, OHF Elman neural network was first selected for training and testing based on the sample data. The results are: the averaged absolute error is 0.0511, the maximum absolute error is 2, and the minimum absolute error is 8.4272e-19, thus τ is determined to be 0.01.

- Determination of the number of base learners: l

The superimposed effect of the weak learners determines the learning ability of the final strong learners. However, with the increase of the number of weak learners, it is uncertain whether the diagnosis effect will continue to be improved or even causing overfitting. Considering the iteration time and effectiveness of ensemble learning, the number range of weak learners is set as [8,15]. The corresponding experimental results are shown in Fig. 8.

To select the number of weak learners, the sum of the MSE errors of all test samples is used as the criterion, as shown in the vertical axis in Fig. 8. It is illustrated that the number of weak learners has a non-linear effect on the strong learners, and the sum of the diagnostic errors takes the minimum when the number of weak learners is set as 13. Thus, the final number of weak learners is $l = 13$.

5.3. Parameters determination of Bagging

The main content of Bagging ensemble is to construct different training datasets through the sampling method. Each dataset is repeatedly sampled from the original dataset. Therefore, each data set lacks some samples from the original data set with high probability and contains several repeated samples. Different training datasets lead to the differences between the training-based models, i.e., OHF Elman neural network.

In Bagging ensemble, two parameters need to be optimized: the number of samples in the training datasets to be constructed, and the number of weak learners. To ensure the consistency with the training set information in other methods, the number of the samples in the training subset n_0 is selected as 550. As for the number

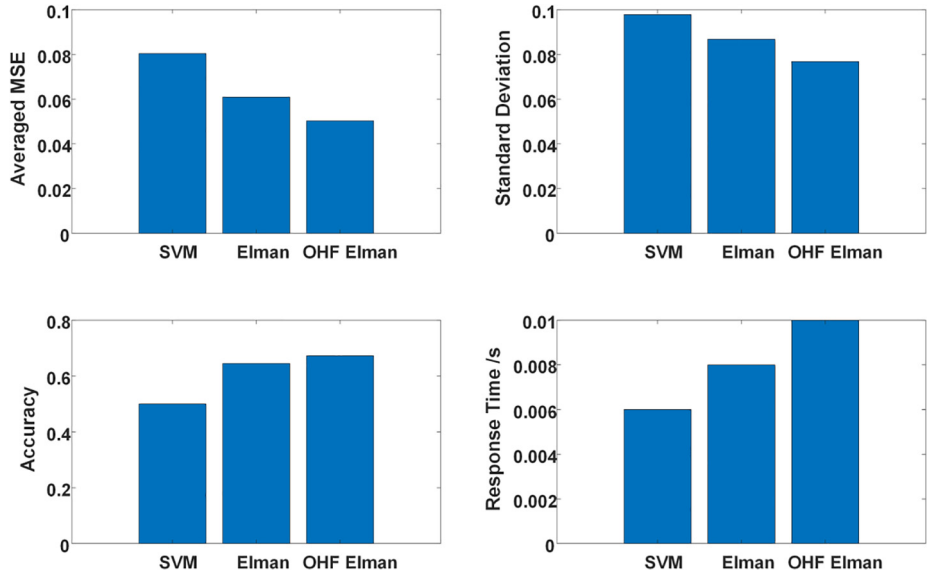


Fig. 7. Diagnostic results of SVM and two neural networks over the entire test dataset.

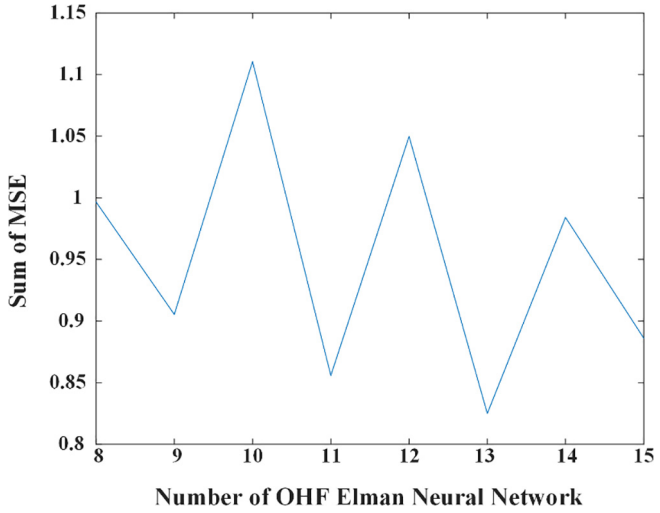


Fig. 8. Diagnostic errors of AdaBoost ensemble with different numbers of weak learners.

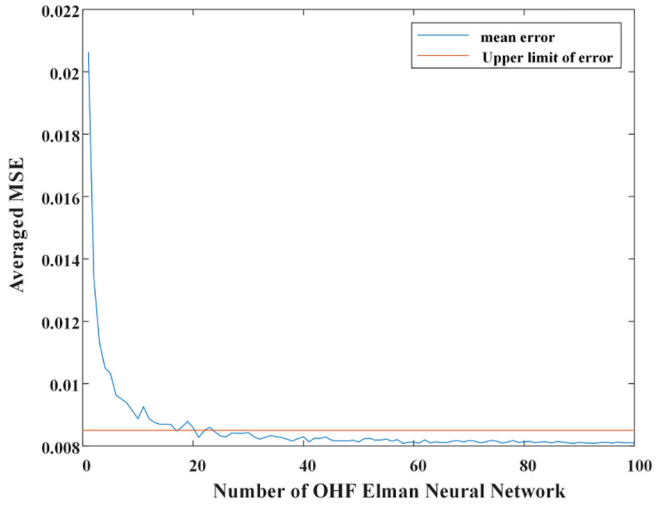


Fig. 9. Diagnostic errors of Bagging ensemble with different numbers of weak learners.

of weak learners, the diagnostic error results of the ensemble models under different numbers of weak learners are shown in Fig. 9.

Here, 100 OHF Elman neural networks are first trained as the base models. Then, K base models are randomly selected and averaged. This process is repeated 30 times, and the averaged result of 30 times is used as the diagnostic error of strong learner under K weak learners. In this process, K is traversed from 1 to 100.

It can be found that the averaged MSE decreases significantly in the early stage of the increase in the number of weak learners. The diagnostic error of the strong learner under 100 weak learners is 0.0081. In this study, the upper limit of the averaged MSE is set to 0.0085 by considering the iteration time and accuracy. It can be found from Fig. 9 that 20 is more suitable as the number of weak learners for Bagging Ensemble.

5.4. Results analysis of different ensemble algorithms

To verify the effectiveness and superiority of the proposed OHF Elman AdaBoost-Bagging Algorithm for rolling bearing fault diagnosis, the same test dataset as OHF Elman neural network, Bagging ensemble and AdaBoost ensemble is selected. And the generalization error and standard deviation are used as the evaluation indicators. Then, the diagnostic effect of each ensemble model from the perspective of each test sample to the entire test dataset will be analyzed. The specific results of OHF Elman AdaBoost-Bagging Algorithm and other two ensemble algorithms are shown in Fig. 10.

Fig. 10(a) shows the diagnostic error MSE on each test sample of OHF Elman neural network, AdaBoost ensemble, Bagging ensemble, and our OHF Elman AdaBoost-Bagging Algorithm. The improvement effects of AdaBoost ensemble algorithm and Bagging ensemble algorithm on weak learners (OHF Elman neural networks) are illustrated in Fig. 10(b)–(d) describes the diagnostic results of AdaBoost ensemble algorithm, Bagging ensemble and AdaBoost-Bagging ensemble algorithm.

From Fig. 10(b) and (c), it can be seen that both Bagging ensemble and AdaBoost ensemble have a very good optimization on the OHF Elman neural network. Bagging ensemble effectively reduces

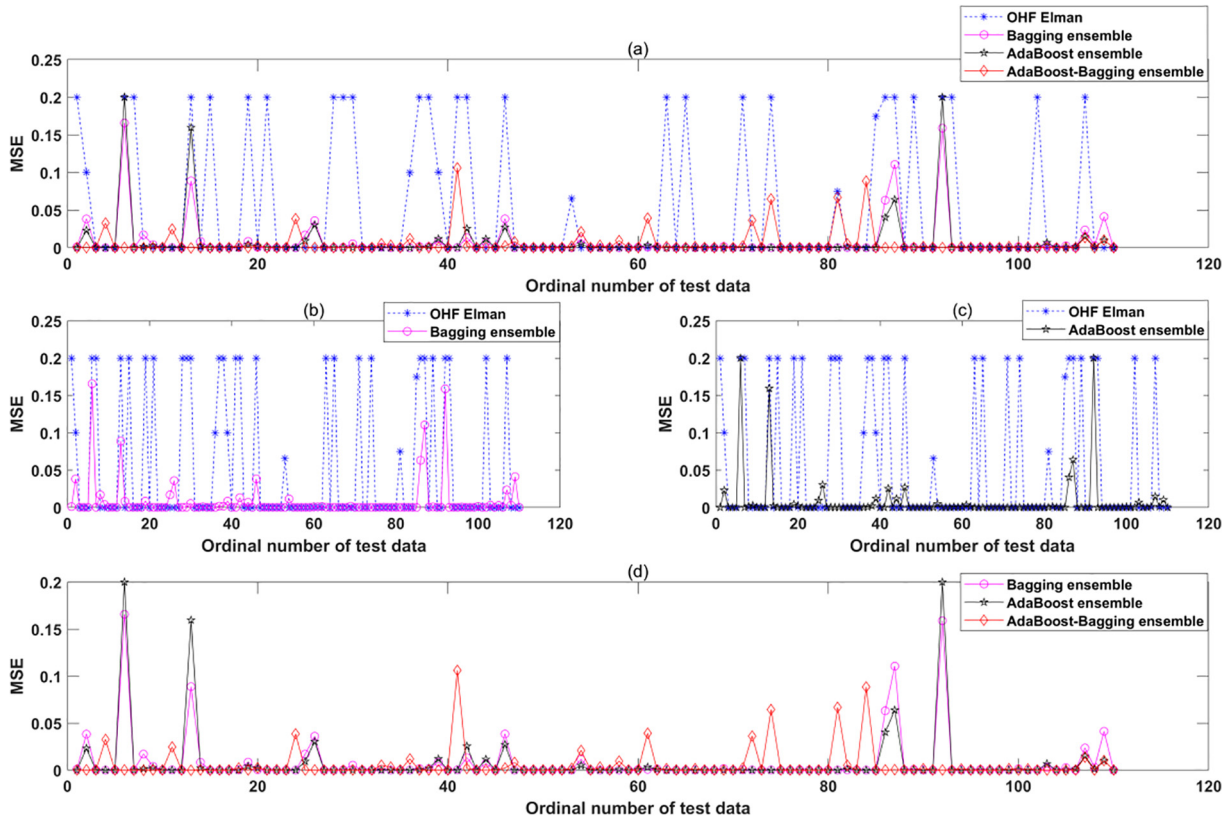


Fig. 10. Diagnostic errors of different algorithms on each test sample.

the extremum of diagnostic errors, which also illustrates the function of Bagging ensemble in improving the stability of the algorithm. In Fig. 10(d), AdaBoost-Bagging ensemble makes full use of the advantages of AdaBoost ensemble and Bagging ensemble. It not only effectively balances the diagnostic effect of each test sample, but also reduces the diagnostic error of the entire test dataset.

The diagnosis results of the above algorithms are different on various samples. To better measure the superiority of the algorithm, we use the accuracy, response time and averaged MSE and standard deviation as the performance metrics. Performance metric values of OHF Elman neural network and three ensemble algorithms are illustrated in Table 7 and Fig. 11.

When Neural Network is performed for rolling bearing fault diagnosis, the impact of anomalous samples on the network model cannot be resolved. Here, we consider that AdaBoost ensemble has more advantages in improving the accuracy of fault diagnosis, and Bagging ensemble can improve the stability of diagnosis model. Thus, Fig. 11 shows that AdaBoost-Bagging ensemble effectively reduces both the averaged MSE and the standard deviation through the dual integration of AdaBoost and Bagging. Although the ensemble model further increases the response time, the response time of 1.6 s can already meet the actual industrial needs.

5.5. Results analysis of multiple faults

When the rolling bearing is in the incipient stage of failure, its vibration signal is abnormal and inconspicuous, resulting in fuzzy fault information. Hence, due to the lack of control over the degree of faults, the abnormal state of rolling bearing is often discovered in the late fault stage, and the delay in fault diagnosis will cause serious failure risk.

In order to effectively control the degree of faults, we classify the rolling bearing faults into incipient, intermediate and late stages according to the fault diameter of single-point. Therefore, in addition to the fault diagnosis of different parts (the inner race, the outer race and the ball), we also study the diagnosis for different fault stages. Tables 8 and 9 describe the specific diagnosis errors of different algorithms for different fault stages and locations. The fault diagnosis results are shown in Fig. 12.

As can be seen from Fig. 12, the diagnostic accuracy of OHF Elman neural network for the normal state is high, which is precisely because the normal signal is relatively stable. For the fault signal prone to be anomalous sample, OHF Elman neural network has a higher diagnostic error. Meanwhile, all algorithms have their highest diagnostic error for the middle stage of faults. Although OHF Elman-AdaBoost has lower diagnostic errors for the normal

Table 7
Diagnostic results of different algorithms on the entire test dataset.

Performance Metrics	OHF Elman	Bagging ensemble	AdaBoost ensemble	AdaBoost-Bagging ensemble
Accuracy	67.27%	94.55%	93.64%	95.45%
Averaged MSE	0.0503	0.0081	0.0078	0.0056
Standard Deviation	0.0768	0.0266	0.0315	0.0170
Response Time/s	0.0100	0.2030	0.1090	1.6570

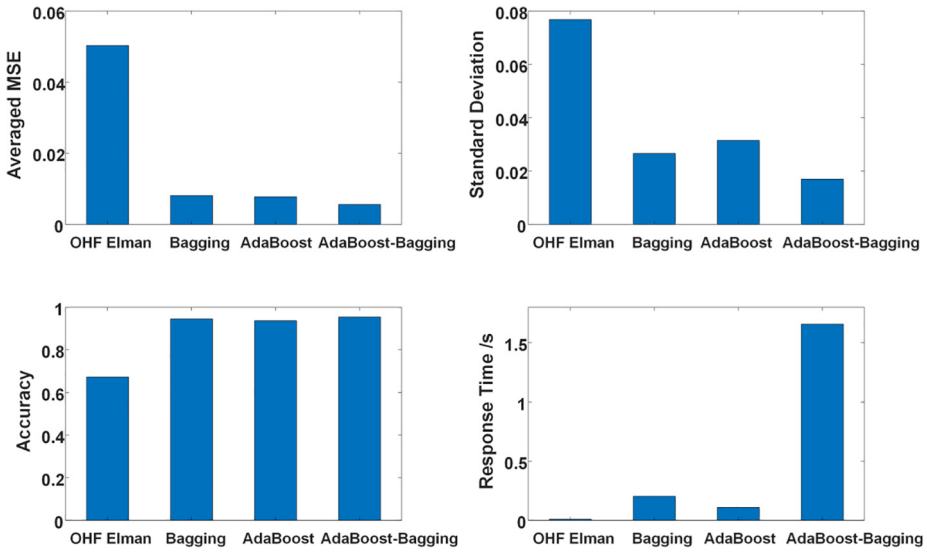


Fig. 11. Results of OHF Elman neural network and three ensemble algorithms.

Table 8
Diagnostic errors (averaged MSE) of algorithms in different fault stages.

	OHF Elman	OHF Elman-Bagging	OHF Elman-AdaBoost	OHF Elman-AdaBoost-Bagging
Normal	1.5688e−12	0.0048	0.0022	0.0031
Incipient	0.0637	1.1425e−04	1.2740e−05	9.8902e−05
Intermediate	0.0983	0.0197	0.0203	0.0138
Late	0.0165	0.0011	0.0014	0.0073

Table 9
Diagnostic errors (averaged MSE) of algorithms in different fault locations.

	OHF Elman	OHF Elman-Bagging	OHF Elman-AdaBoost	OHF Elman-AdaBoost-Bagging
Normal	1.5688e−12	0.0048	0.0022	0.0031
Ball	0.0532	0.0040	0.0036	0.0138
Inner race	0.0118	5.3720e−04	2.3275e−04	2.4511e−04
Outer race	0.1068	0.0181	0.0194	0.0035

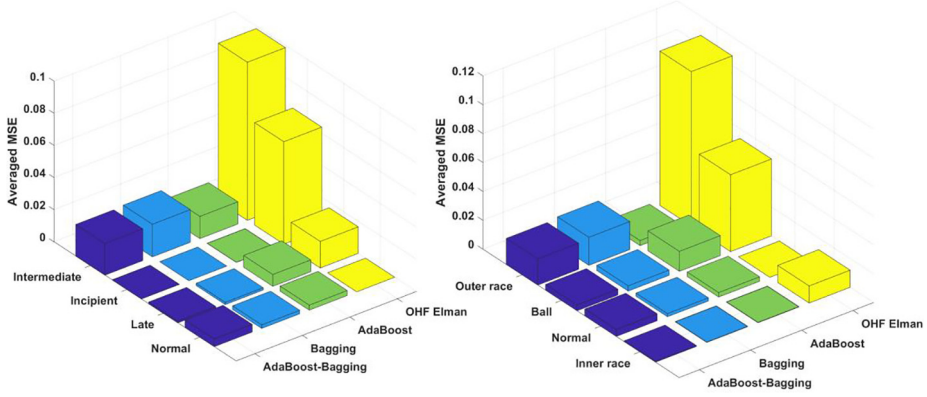


Fig. 12. Diagnostic errors of different algorithms in different fault stages and locations.

state, the incipient stage and the late stage of faults, OHF Elman-AdaBoost and OHF Elman AdaBoost-Bagging algorithms have little difference in the diagnosis effect due to their lower magnitude. Instead, effective fault diagnosis of the intermediate stage can arrange reasonable maintenance and replacement strategies. Therefore, the proposed OHF Elman AdaBoost-Bagging algorithm, which has lower diagnostic errors for the intermediate stage of faults, has more advantages.

Besides, all the algorithms have higher fault diagnosis errors for the ball or the outer race. AdaBoost ensemble and Bagging ensemble have greatly improved the fault diagnosis accuracy of all locations, but the fault diagnosis error of the outer race is the largest.

Table 10
Diagnostic errors (averaged MSE) for various locations in different stages.

	Ball	Inner race	Outer race
Incipient	3.0033e−04	3.7405e−06	7.5469e−06
Intermediate	0.0264	7.6308e−04	0.0129
Late	0.0200	5.0945e−05	2.9532e−04

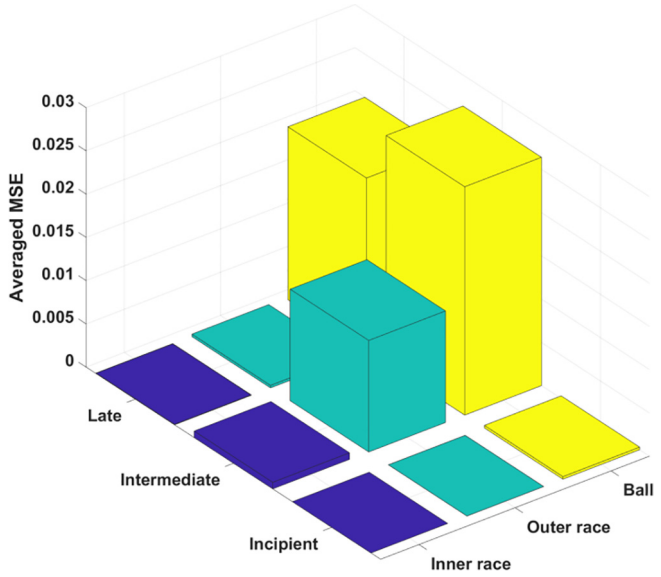


Fig. 13. Diagnostic errors of OHF Elman AdaBoost-Bagging algorithm.

Although the OHF Elman AdaBoost-Bagging algorithm improves the accuracy of fault diagnosis for the outer ring, the fault diagnosis of rolling elements has a large error. In view of this problem, the diagnostic errors of the OHF Elman AdaBoost-Bagging algorithm are further analyzed for different fault stages of various locations, as shown in Table 10 and Fig. 13.

Fig. 13 describes the fault diagnosis results of the OHF Elman AdaBoost-Bagging algorithm for various locations in different stages. It can be found that the diagnostic errors of the OHF Elman AdaBoost-Bagging algorithm for the ball mainly come from the intermediate and late stages. Besides, the OHF Elman AdaBoost-Bagging algorithm has the good diagnostic accuracy for the late stage of the outer race and the inner race.

6. Conclusion

In this paper, we propose an OHF Elman AdaBoost-Bagging algorithm for multi-stage fault diagnosis of rolling bearing, which aims to improve the diagnostic accuracy and stability through the dual integration of AdaBoost ensemble and Bagging ensemble. First of all, the signal decomposition and reconstruction are performed by using EEMD, and six valid features are extracted. Then, the multi-stage (incipient, intermediate, late) fault diagnosis under random noise is achieved for different parts/locations (the inner race, the outer race and the ball) and the OHF Elman AdaBoost-Bagging algorithm is verified to be more suitable for multi-stage fault diagnosis of rolling bearings, which has achieved 95.45% classification accuracy and a lower diagnostic error as 0.0056. The effectiveness of the feedback mechanism of OHF Elman neural network is illustrated by the comparison experiment of Elman neural network, OHF Elman neural network and SVM. Through further analysis of diagnostic errors (averaged MSE) for various locations in different stages, the diagnostic errors of the OHF Elman AdaBoost-Bagging algorithm for the ball mainly come from the intermediate and late stages.

Although this OHF Elman AdaBoost-Bagging algorithm has achieved a promising performance, there is still ample space for further improvements, mainly to further improve the diagnostic accuracy of rolling bearing faults in the intermediate stage and the ball part. The future work is to extract more effective features through more detailed data processing. In addition, parameter optimization of the OHF Elman AdaBoost-Bagging algorithm will be carried out by other methods to further improve diagnostic accuracy.

CRediT authorship contribution statement

Tangbin Xia: Conceptualization, Methodology, Writing - original draft. **Pengcheng Zhuo:** Methodology, Writing - original draft, Software. **Lei Xiao:** . **Shichang Du:** Supervision, Investigation. **Dong Wang:** Software, Validation. **Lifeng Xi:** .

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research is funded by National Natural Science Foundation of China (51875359, 51705321, 51975355), Natural Science Foundation of Shanghai (20ZR1428600) and Ministry of Education-China Mobile Research Foundation (CMHQ-JS-201900003).

References

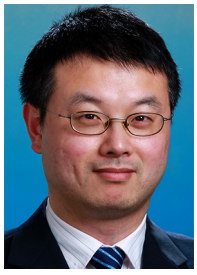
[1] Y. Lei, Z. Qiao, X. Xu, J. Lin, S. Niu, An underdamped stochastic resonance method with stable-state matching for incipient fault diagnosis of rolling element bearings, Mech. Syst. Sig. Process. 94 (2017) 148–164, https://doi.org/10.1016/j.ymssp.2017.02.041.
[2] X. Li, W. Zhang, Q. Ding, A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning, Neurocomputing 310 (2018) 77–95, https://doi.org/10.1016/j.neucom.2018.05.021.
[3] W. Huang, H. Sun, J. Luo, W. Wang, Periodic feature oriented adapted dictionary free OMP for rolling element bearing incipient fault diagnosis, Mech. Syst. Sig. Process. 126 (2019) 137–160, https://doi.org/10.1016/j.ymssp.2019.02.023.
[4] C. He, P. Niu, R. Yang, C. Wang, Z. Li, H. Li, Incipient rolling element bearing weak fault feature extraction based on adaptive second-order stochastic resonance incorporated by mode decomposition, Measurement 145 (2019) 687–701, https://doi.org/10.1016/j.measurement.2019.05.052.
[5] D. Wang, K.-L. Tsui, Two novel mixed effects models for prognostics of rolling element bearings, Mech. Syst. Sig. Process. 99 (2018) 1–13, https://doi.org/10.1016/j.ymssp.2017.06.004.
[6] D. Wang, Q. Miao, Q. Zhou, G. Zhou, An intelligent prognostic system for gear performance degradation assessment and remaining useful life estimation, J. Vib. Acoust. Trans. ASME. 137 (2015), https://doi.org/10.1115/1.4028833 021004.
[7] X. Yan, Y. Liu, M. Jia, Research on an enhanced scale morphological-hat product filtering in incipient fault detection of rolling element bearings, Measurement 147 (2019) 106856, https://doi.org/10.1016/j.measurement.2019.106856.
[8] Z. Meng, X. Zhan, J. Li, Z. Pan, An enhancement denoising autoencoder for rolling bearing fault diagnosis, Measurement 130 (2018) 448–454, https://doi.org/10.1016/j.measurement.2018.08.010.
[9] Z. German-Sallo, H.S. Grif, Hilbert-huang transform in fault detection, Procedia Manuf. 32 (2019) 591–595, https://doi.org/10.1016/j.promfg.2019.02.257.
[10] P.K. Kankar, S.C. Sharma, S.P. Harsha, Fault diagnosis of rolling element bearing using cyclic autocorrelation and wavelet transform, Neurocomputing 110 (2013) 9–17, https://doi.org/10.1016/j.neucom.2012.11.012.
[11] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Snin, Q. Zheng, N.C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. A Math. Phys. Eng. Sci. 454 (1998) 903–995, https://doi.org/10.1098/rspa.1998.0193.
[12] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: A noise-assisted data analysis method, Adv. Adapt. Data Anal. 01 (01) (2009) 1–41, https://doi.org/10.1142/S1793536909000047.

[13] T. Xia, Y. Dong, L. Xiao, S. Du, E. Pan, L. Xi, Recent advances in prognostics and health management for advanced manufacturing paradigms, Reliab. Eng. Syst. Saf. 178 (2018) 255–268, https://doi.org/10.1016/j.ress.2018.06.021.
[14] T. Xia, X. Fang, N. Gebraeel, L. Xi, E. Pan, Online analytics framework of sensor-driven prognosis and opportunistic maintenance for mass customization, J. Manuf. Sci. Eng. Trans. ASME. 141 (2019), https://doi.org/10.1115/1.4043255 051011.
[15] Y. Dong, T. Xia, X. Fang, Z. Zhang, L. Xi, Prognostic and health management for adaptive manufacturing systems with online sensors and flexible structures, Comput. Ind. Eng. 133 (2019) 57–68, https://doi.org/10.1016/j.cie.2019.04.051.
[16] T. Xia, X. Jin, L. Xi, Y. Zhang, J. Ni, Operating load based real-time rolling grey forecasting for machine health prognosis in dynamic maintenance schedule, J. Intell. Manuf. 26 (2) (2015) 269–280, https://doi.org/10.1007/s10845-013-0780-8.
[17] J. Li, X. Zhang, X. Zhou, L. Lu, Reliability assessment of wind turbine bearing based on the degradation-Hidden-Markov model, Renewable Energy 132 (2019) 1076–1087, https://doi.org/10.1016/j.renene.2018.08.048.
[18] X. Yan, M. Jia, A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing, Neurocomputing 313 (2018) 47–64, https://doi.org/10.1016/j.neucom.2018.05.002.
[19] W. Lu, Y. Li, Y.u. Cheng, D. Meng, B. Liang, P. Zhou, Early fault detection approach with deep architectures, IEEE Trans. Instrum. Meas. 67 (7) (2018) 1679–1689, https://doi.org/10.1109/TIM.2018.2800978.
[20] Y. Li, Y. Yang, X. Wang, B. Liu, X. Liang, Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine, J. Sound Vib. 428 (2018) 72–86, https://doi.org/10.1016/j.jsv.2018.04.036.
[21] P.R.N. da Silva, H.A. Gabbar, P. Vieira Junior, C.T. da Costa Junior, A new methodology for multiple incipient fault diagnosis in transmission lines using QTA and Naïve Bayes classifier, Int. J. Electr. Power Energy Syst. 103 (2018) 326–346, https://doi.org/10.1016/j.ijepes.2018.05.036.
[22] M.S. Hossain, Z.C. Ong, Z. Ismail, S. Noroozi, S.Y. Khoo, Artificial neural networks for vibration based inverse parametric identifications: A review, Appl. Soft Comput. 52 (2017) 203–219, https://doi.org/10.1016/j.asoc.2016.12.014.
[23] L. Yang, F. Wang, J. Zhang, W. Ren, Remaining useful life prediction of ultrasonic motor based on Elman neural network with improved particle swarm optimization, Measurement 143 (2019) 27–38, https://doi.org/10.1016/j.measurement.2019.05.013.
[24] T. Xia, Y.a. Song, Y.u. Zheng, E. Pan, L. Xi, An ensemble framework based on convolutional bi-directional LSTM with multiple time windows for remaining useful life estimation, Comput. Ind. 115 (2020) 103182, https://doi.org/10.1016/j.compind.2019.103182.
[25] W. Huang, J. Cheng, Y.u. Yang, G. Guo, An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis, Neurocomputing 359 (2019) 77–92, https://doi.org/10.1016/j.neucom.2019.05.052.
[26] X.H. Shi, Y.C. Liang, H.P. Lee, W.Z. Lin, X. Xu, S.P. Lim, Improved elman networks and applications for controlling ultrasonic motors, Appl. Artif. Intell. 18 (7) (2004) 603–629, https://doi.org/10.1080/08839510490483279.
[27] B.S. Raghuwanshi, S. Shukla, Class-specific extreme learning machine for handling binary class imbalance problem, Neural Networks 105 (2018) 206–217, https://doi.org/10.1016/j.neunet.2018.05.011.
[28] T. Takenouchi, S. Ishii, Binary classifiers ensemble based on Bregman divergence for multi-class classification, Neurocomputing 273 (2018) 424–434, https://doi.org/10.1016/j.neucom.2017.08.004.
[29] N. García-Pedrajas, D. Ortiz-Boyer, An empirical study of binary classifier fusion methods for multiclass classification, Inf. Fusion 12 (2) (2011) 111–130, https://doi.org/10.1016/j.inffus.2010.06.010.
[30] M.H.D.M. Ribeiro, L. dos Santos Coelho, Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series, Appl. Soft Comput. 86 (2020) 105837, https://doi.org/10.1016/j.asoc.2019.105837.
[31] X. Zhang, B. Wang, X. Chen, Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine, Knowledge-Based Syst. 89 (2015) 56–85, https://doi.org/10.1016/j.knosys.2015.06.017.
[32] Y. Wu, Y. Ke, Z. Chen, S. Liang, H. Zhao, H. Hong, Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping, Catena 187 (2020), https://doi.org/10.1016/j.catena.2019.104396 104396.
[33] J. Lin, H. Chen, S. Li, Y. Liu, X. Li, B. Yu, Accurate prediction of potential druggable proteins based on genetic algorithm and Bagging-SVM ensemble classifier, Artif. Intell. Med. 98 (2019) 35–47, https://doi.org/10.1016/j.artmed.2019.07.005.
[34] L. Li, Y. Huang, J. Tao, C. Liu, K. Li, Featured temporal segmentation method and AdaBoost-BP detector for internal leakage evaluation of a hydraulic cylinder, Meas. J. Int. Meas. Confed. 130 (2018) 279–289, https://doi.org/10.1016/j.measurement.2018.08.029.
[35] H. Li, T. Liu, X. Wu, Q. Chen, Application of EEMD and improved frequency band entropy in bearing fault feature extraction, ISA Trans. 88 (2019) 170–185, https://doi.org/10.1016/j.isatra.2018.12.002.
[36] L. Breiman, Bagging predictors, Mach. Learn. 24 (1996) 123–140, https://doi.org/10.1007/BF00058655.
[37] S. Moral-García, C.J. Mantas, J.G. Castellano, M.D. Benítez, J. Abellán, Bagging of credal decision trees for imprecise classification, Expert Syst. Appl. 141 (2020), https://doi.org/10.1016/j.eswa.2019.112944.
[38] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1997) 119–139, https://doi.org/10.1006/jcss.1997.1504.
[39] Y. Freund, R.E. Schapire, Experiments with a New Boosting Algorithm, in, in: Proceedings of the 13th International Conference on Machine Learning, 1996, pp. 148–156.
[40] H. Drucker, Improving regressors using boosting techniques, in, in: Proceedings of the 14th International Conference on Machine Learning, 1997, pp. 107–115.
[41] D. Pardoe, P. Stone, Boosting for regression transfer, in, in: Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 863–870.
[42] D.L. Shrestha, D.P. Solomatine, Experiments with AdaBoost.RT, an improved boosting scheme for regression, Neural Comput. 18 (2006) 1678–1710, https://doi.org/10.1162/neco.2006.18.7.1678.
[43] Z. Zhu, G. Peng, Y. Chen, H. Gao, A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis, Neurocomputing 323 (2019) 62–75, https://doi.org/10.1016/j.neucom.2018.09.050.
[44] J. Sun, H. Fujita, P. Chen, H. Li, Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble, Knowledge-Based Syst. 120 (2017) 4–14, https://doi.org/10.1016/j.knosys.2016.12.019.
[45] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study, Mech. Syst. Signal Process. 64–65 (2015) 100–131, https://doi.org/10.1016/j.ymssp.2015.04.021.
[46] S.K. Laha, Enhancement of fault diagnosis of rolling element bearing using maximum kurtosis fast nonlocal means denoising, Meas. J. Int. Meas. Confed. 100 (2017) 157–163, https://doi.org/10.1016/j.measurement.2016.12.058.
[47] P. Kundu, A.K. Darpe, M.S. Kulkarni, A correlation coefficient based vibration indicator for detecting natural pitting progression in spur gears, Mech. Syst. Signal Process. 129 (2019) 741–763, https://doi.org/10.1016/j.ymssp.2019.04.058.
[48] S. Toker, S. Kaçiranlar, On the performance of two parameter ridge estimator under the mean square error criterion, Appl. Math. Comput. 219 (2013) 4718–4728, https://doi.org/10.1016/j.amc.2012.10.088.
[49] X. Zhao, N. Jiang, J. Liu, D. Yu, J. Chang, Short-term average wind speed and turbulent standard deviation forecasts based on one-dimensional convolutional neural network and the integrate method for probabilistic framework, Energy Convers. Manag. 203 (2020), https://doi.org/10.1016/j.enconman.2019.112239 112239.
[50] J. Wang, W. Zhu, W. Zhang, D. Sun, A trend fixed on firstly and seasonal adjustment model combined with the ε-SVR for short-term forecasting of electricity demand, Energy Policy 37 (2009) 4901–4909, https://doi.org/10.1016/j.enpol.2009.06.046.

Tangbin Xia received Ph.D. degree in Mechanical Engineering (Industrial Engineering) from Shanghai Jiao Tong University in 2014 and was a co-cultured Ph.D. with University of Michigan. Currently, he is an Associate Professor and serves as Deputy Director of the Department of Industrial Engineering at Shanghai Jiao Tong University. His major research interests are intelligent maintenance systems, prognostics & health management, and advanced manufacturing. He is a member of IEEE, ASME, IISE and INFORMS.

Pengcheng Zhuo received B.S degree in Industrial Engineering and Management from Shanghai Jiao Tong University, Shanghai, China in 2018. He is currently pursuing the M.S. degree in Industrial Engineering from Shanghai Jiao Tong University, Shanghai, China. His current research interests include data-driven prognostic and health management.

Lei Xiao received Ph.D. degree in Management Science and Engineering from Chongqing University in 2016. She was co-cultured Ph.D. with Rutgers University. She worked as a postdoc at Shanghai Jiao Tong University. Currently, she is an Assistant Professor with the Department of Mechanical Engineering, Donghua University. Her research interests involve the key techniques in predictive maintenance, including fault detection, RUL prediction and maintenance strategies optimization.



Shichang Du received B.S. and M.S.E. degrees in Mechanical Engineering from the Hefei University of Technology, Hefei, China, in 2000 and 2003, respectively, and Ph.D. degree in Industrial Engineering and Management from Shanghai Jiao Tong University, Shanghai, China, in 2008. He was a Visiting Scholar with the University of Michigan. He is a Professor with Shanghai Jiao Tong University. His current research interests include quality and reliability engineering, quality control with analysis of error flow, and monitoring and diagnosis of manufacturing process.

Dong Wang received Ph.D. degree from the City University of Hong Kong, in 2015. Currently, he is an Associate Professor with the Department of Industrial Engineering at Shanghai Jiao Tong University. He has been awarded State Specially Recruited Experts (Young Talents). He is an Associate Editor for IEEE Transactions on Instrumentation and Measurement. His research interests include condition monitoring, fault diagnosis, prognosis and health management, signal processing, data mining, nondestructive testing, and statistical modeling.

Lifeng Xi received Ph.D. degree in Mechanical Engineering from Shanghai Jiao Tong University in 1995. He is currently a Professor and serves as the Vice President of Shanghai Jiao Tong University. He also serves as the Deputy Editor of Industrial Engineering & Management, the Fellow of State Key Laboratory of Mechanical System and Vibration, and the Executive Director of the Chinese Quality Association. He has published more than 200 papers in international journals. His interest areas are in the fields of production system design and planning, quality management, and reliability engineering.