

## Prediction Method of Bearing RUL Based on DNN and GBDT Algorithm

Qu Jinglei<sup>1</sup>, Li Changsheng<sup>1\*</sup>, Chen Weixing<sup>2</sup>, Zhang Chenglong<sup>3</sup>

<sup>1</sup>*School of Mechanical Engineering, Zhengzhou Henan Institute of Technology, Xinxiang, Henan, China*

<sup>2</sup>*Hangzhou Eohi Cyber System Co., Ltd., Hangzhou, China*

<sup>3</sup>*School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China*

*\*Corresponding Author.*

### Abstract

*In view of the lack of effective model and the large prediction error in the traditional prediction methods, a collaborative prediction method for remaining useful life of bearing based on DNN and GBDT is proposed. Firstly, the degradation characteristics are constructed through normalization processing of parameters in time domain and frequency domain that can clearly represent the healthy running state of bearings, in order to improve the correlation of degradation characteristics, the prior model features are generated based on DNN. Secondly, a regression model of GBDT based on the prior model features is presented. Finally, the experimental results show that compared with other algorithms such as DNN, GBDT, SVR, RF, DT, the proposed method has better prediction performance evaluation results, higher prediction accuracy and efficiency compared with other algorithms.*

**Keywords:** *remaining useful life prediction, bearing, gradient boosting decision tree, health management, deep neural network*

### I. Introduction

With the integration of information technology and manufacturing industry more and more extensive and in-depth, the mechanical equipment is becoming more and more complex, accurate and intelligent. With the continuous operation of mechanical equipment, its running state and key parts will gradually degenerate, and the probability of failure and shutdown will gradually increase, which will affect the normal production and processing of enterprises. Therefore, the service life prediction and health management of mechanical equipment have attracted wide attention. It is of great significance to improve the safety and reliability of key mechanical equipment by collecting, analyzing and processing massive monitoring data of mechanical equipment running condition, judging the operation status of equipment, and then carrying out preventive maintenance of equipment<sup>[1]</sup>. As an important part of mechanical equipment, bearing is one of the parts with high probability of failure, which usually runs in high-speed and heavy load conditions. Therefore, it is often used as a typical part for equipment fault prediction and health management. The traditional methods for prediction and analysis methods of bearing remaining useful life (RUL) are mainly divided into physical model-based, statistical analysis based and knowledge-based method. However, due to the lack of effective models and strong dependence on historical experience, the prediction results often deviate greatly from the actual operation of the equipment.

Artificial intelligence, big data and other technologies have gained rapid development and wide application in recent years, while data-driven bearing RUL prediction has developed into an active research direction due to its strong adaptability. Through machine learning, deep learning and other methods, the feature extraction, modeling and analysis are carried out on the historical running state data of equipment, which has high prediction accuracy for bearing residual life prediction. In reference [2], an adaptive multi-core combined relevance vector machine is proposed to predict the RUL of electromechanical equipment. In the literature [3], aiming at the problem of bearing fault prediction and health management, a deep neural networks (DNN) learning algorithm based on sparse self

encoding is proposed. In reference [4], an 18 dimensional degradation characteristic index based on bearing vibration signal is established, and a predictive algorithm based on deep learning is designed. Wei et al. designed a trend analysis method for vibration characteristics based on the autoregressive moving average (ARIMA) in [5]. In order to achieve higher prediction accuracy and effect, Liu et al. proposed to use quadratic programming to improve the robustness<sup>[6]</sup>. In addition, Wu et al. used random forest regression method to predict tool wear during milling<sup>[7]</sup>. With the rapid improvement of computing power and computational efficiency, deep learning approach has become one of the hot research topics in the field of prediction because of its powerful learning ability in the complex system<sup>[8]</sup>. In most cases, deep learning uses vibration signal to monitor the bearing health, because the signal contains important information about the occurrence of faults<sup>[9]</sup>. Hasani et al. in [10] in order to accurately identify the starting point of bearing degradation, a prediction algorithm based on automatic encoder was designed. Wang et al. collected the data sample of polishing materials and proposed a prediction method of polishing material removal rate based on depth belief network (DBN)<sup>[11]</sup>. Liao et al. calculated the operation data of the machine and proposed a Boltzmann machine (RBM) with regularization term and proved the prediction effect by experiments<sup>[12]</sup>. The artificial intelligence method has higher prediction accuracy than traditional bearing RUL prediction method, but it has the problems of large prediction error and low calculation efficiency<sup>[13]</sup>.

Therefore, to improve the correlation of degradation features, increase the prediction accuracy and improve the prediction efficiency of the model, the DNN and the gradient boosting decision tree (GBDT) model are combined to ensure that the model has a strong feature learning ability and improve model generalization ability. This paper propose a bearing RUL collaborative prediction method based on DNN and GBDT. The root mean square (RMS), peak factor, kurtosis factor and frequency spectrum partition summation of the vibration signal are selected as degradation characteristic indexes. The experimental results show that, compared with GBDT, DNN, support vector regression (SVR), random forests (RF), decision tree (DT), Gaussian process (GP) and other models, the proposed method is more effective, it has high prediction accuracy.

## II. Construction of Bearing Degradation Characteristics

The traditional method to extract the bearing degradation characteristics is mainly to do spectrum analysis on bearing vibration signals, and generally extract various statistics as features in time , frequency and time-frequency domain<sup>[14]</sup>. The prediction effect is often disturbed by a great deal of random noise in the traditional rolling bearing life prediction model. By selecting the characteristic index of bearing degradation with obvious characteristics and carrying out fitting analysis with its historical operating state data and error evaluation, the RUL prediction value can be more accurate .

In order to overcome the influence of the numerical range among different characteristic parameters, reduce the data dimension and improve the algorithm efficiency, this paper selects the parameters that can obviously represent the healthy running state of the bearing in the time and frequency domain characteristics of the bearing vibration signals from horizontal and vertical directions respectively, which are Eq. (1) root mean square, Eq. (2) peak factor, Eq. (3) kurtosis factor and Eq. (4) frequency spectrum partition summation (FSPS). To solve the interference of different data dimensions, the maximum and minimum normalization method is used for eigenvalue processing, as shown in Eq. (5). Finally, the 18 dimensional degradation characteristic index based on the bearing vibration signal is constructed, see Eq. (6).

$$X_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (1)$$

$$X_{crest} = \frac{\max(x_i)}{X_{RMS}} \quad (2)$$

$$X_{kurtosis} = \frac{1}{N} \frac{\sum_{i=1}^N (|x_i - \bar{x}|)^4}{X_{RMS}^4} \quad (3)$$

$$X_{FSPS}(m) = \sum_{k=\frac{M+N_{\beta}(k-1)}{M}}^{\frac{N_{\beta}m}{M}} S(k) \quad (4)$$

Among them, FSPS is a one-dimensional frequency domain feature composed of  $M$  elements,  $m$  is the element number of FSPS feature component, where  $m=1,2,3,\dots,M$ ;  $M$  is an empirically determined parameter value, which is set according to the specific problem, and the value in this paper is 6;  $N_{\beta}$  is the spectrum length of the frequency domain signal after Fourier transform;  $k$  is the spectral line numbers for each spectrum;  $S(k)$  is the spectrum value of the input signal.

$$f_j^* = \frac{f_j - f_{j\min}}{f_{j\max} - f_{j\min}} \quad (5)$$

Where  $j$  represents different characteristics of bearing vibration signals;  $f_j$  represents the value of  $j$  of the original bearing vibration signal;  $f_{j\min}$  is the minimum value of  $f_j$ ;  $f_{j\max}$  is the maximum value of  $f_j$ ;  $f_j^*$  represents the normalized value of characteristic  $j$ .

$$[X_i^j] = (x_i^1, x_i^2, x_i^3, \dots, x_i^{16}, x_i^{17}, x_i^{18}) \quad (6)$$

Where  $i$  represents the data point;  $x_i^1 - x_i^3$  represents the root mean square, peak factor and kurtosis factor of the horizontal vibration signal;  $x_i^4 - x_i^9$  represents 6-dimensional FSPS frequency domain feature in the horizontal direction;  $x_i^{10} - x_i^{12}$  represents Represent three time domain characteristics of bearing vibration signal in vertical direction, namely root mean square, peak factor and kurtosis factor;  $x_i^{13} - x_i^{18}$  represents 6-dimensional FSPS frequency domain feature in the vertical direction

### III. Cooperative Prediction Model Based on DNN and GBDT

#### 3.1 Deep neural network

Neural networks (NN) is a mathematical model for information processing by simulating the characteristics of biological neural networks. With the improvement of the demand for training efficiency, Professor Geoffrey Hinton changed the traditional training method of neural networks and improved the training efficiency of multi-layer neural networks. For the first time, he increased the number of network layers to seven and proposed the concept of DNN.

Similar to the traditional artificial neural network structure, DNN including the following parts: an input layer, an  $n$ -layer hidden layer and an output layer, see Fig. 1.  $X$  represents  $m$ -dimension input data, input layer transfers the data to hidden layer 1, performs linear transformation through activation function, and transfers the transformation result to hidden layer 2, and repeatedly runs to output layer repeatedly to obtain the output result  $y$ . The number of DNN network layers can reach up to 20 layers or more according to different application scenarios.

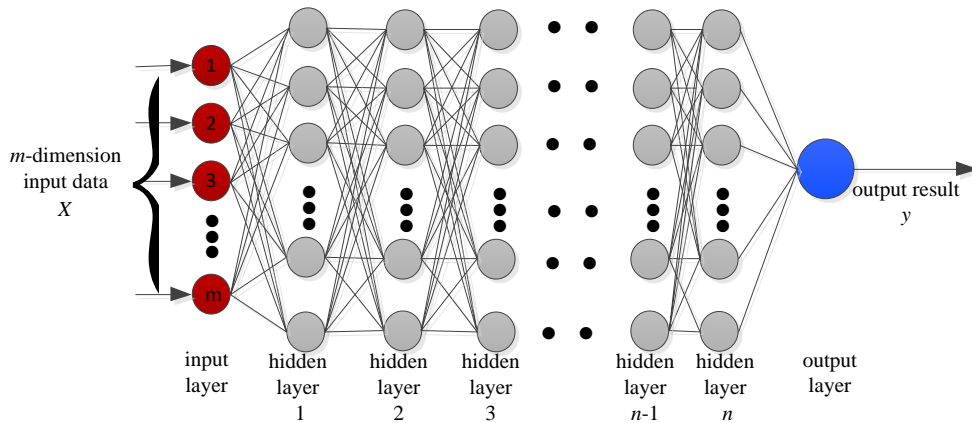


Fig 1: Deep neural network model

In order to avoid over fitting phenomenon in DNN network training process and improve training efficiency, Adam optimization function is optimized in DNN training, and learning rate attenuation mechanism is used to train each DNN network model. In the initial stage of DNN model training, the learning rate is large and the loss function decreases quickly. In the later training stage, the learning rate is small, which can guarantee the model convergence. The calculation formula is as follows:

$$lr = lr_{\min} + (lr_{\max} - lr_{\min}) e^{-\frac{itr}{decay}} \quad (7)$$

Where  $lr_{\max}$  and  $lr_{\min}$  represents the maximum learning rate and the minimum learning rate respectively;  $itr$  represents the number of training iterations of DNN model;  $decay$  represents the decline rate of learning rate.

### 3.2 GBDT

GBDT is an iterative decision algorithm composed of gradient boosting, regression decision tree and shrinking principles. By constructing a plurality of weak classifiers, iterative training is carried out to fit the residual correction model of the previous generation of samples, and finally a strong learner is obtained, which can be effectively used for feature combination and has strong generalization ability. The algorithm flow chart is shown in Fig. 2.

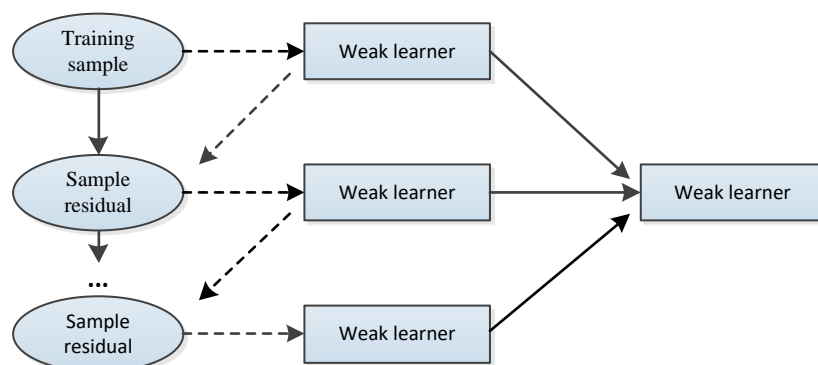


Fig 2: GBDT algorithm flow chart

Taking DNN prior model characteristics as input data X and Y as corresponding label values, GBDT regression process is as follows:

Step 1: Initialize the loss function;  $F_0(X) = \bar{y}$

Step 2: Select Eq. (8) Least-square as the loss function, and calculate the negative gradient value  $\tilde{y}$  of each tree, see Eq. (9), where  $l$  is the number of the tree, and the value range is  $l = 1, 2, 3, \dots, L$ ; the number of decision trees in this article  $L$  is set to 100.

$$L(y_i, F(X_i)) = \frac{1}{2} \times (y_i - F(X_i))^2 \quad (8)$$

$$\begin{aligned} \tilde{y}_i &= -[\partial L(\hat{y}_i) = -\left. \frac{\partial L(y_i, F(X_i))}{\partial F(X_i)} \right]_{F(X)=F_{l-1}(X)} \\ &= (y_i - F_{l-1}(X_i)) \end{aligned} \quad (9)$$

Step 3: Taking the negative gradient value of each tree obtained by Step 2 as training data, regression tree fitting is carried out, as shown in Eq. (10), and leaf node areas are obtained. Among them,  $p = 1, 2, 3, \dots, P$  represents the leaf nodes number, and  $P$  represents the leaf nodes maximum number. In order to avoid over-fitting of the algorithm, the maximum number of leaf nodes is usually not limited.

$$\{R_{pl}\}_1^P = P - \text{terminal node tree}(\{\tilde{y}_i, X_i\}_1^N) \quad (10)$$

Step 4: Superimposing the prediction result of the trained  $l-1$  tree with the  $l$ -th tree, see Eq. (11), where  $\gamma_{pl} = \text{ave}_{X_i \in R_{pl}} \tilde{y}_i$  represents the  $p$ -th leaf node value of the  $l$ -th tree and the average value of the negative gradient value of the sample  $i$ ; *learning\_rate* is the leaf node learning rate, which is set to 0.1 in this article.

$$F_l(X) = F_{l-1}(X) + \text{learning\_rate} \times \sum_{p=1}^P \gamma_{pl} I(X \in R_{pl}) \quad (11)$$

Step 5: Execute Step 2-Step 4 iterative optimization to obtain a regression prediction value  $\hat{y}$  of the bearing RUL .

### 3.3 Collaborative forecasting model based on DNN and GBDT

To effectively combine the feature learning ability of DNN with the feature combination and generalization ability of GBDT, a collaborative prediction model based on DNN and GBDT is designed. The flow chart is shown in Figure 3.

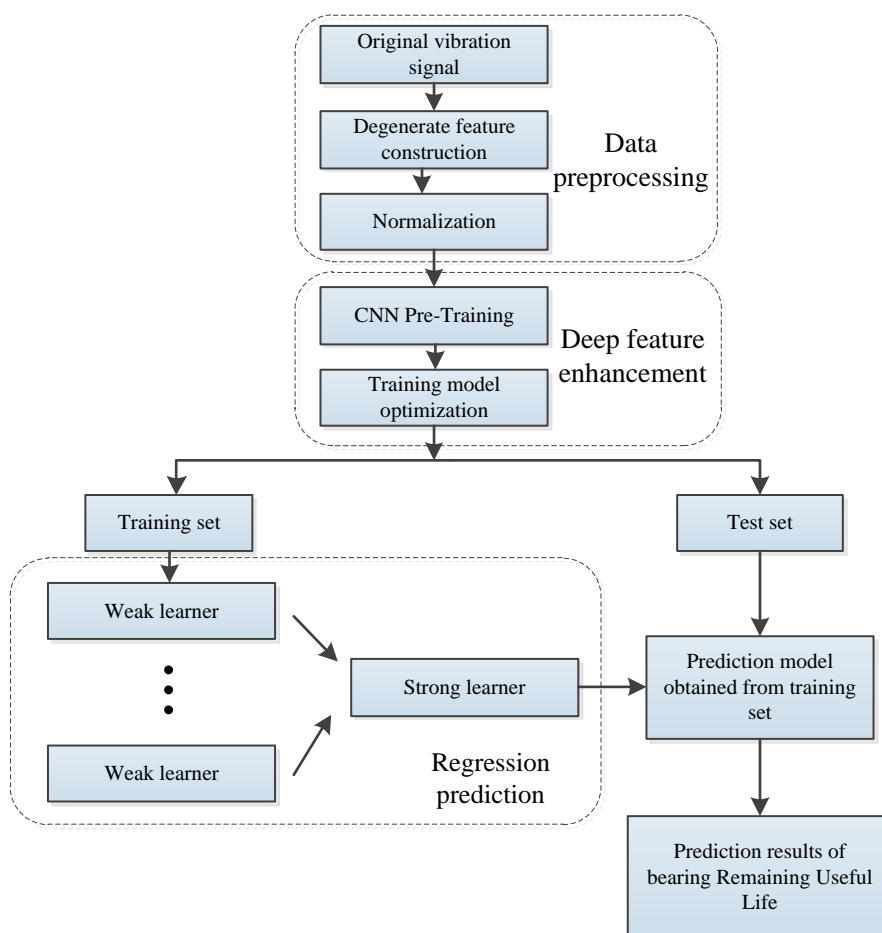


Fig. 3: Flow chart of collaborative prediction model based on DNN and GBDT

(1) Data preprocessing: Firstly, the original vibration signal is preprocessed, the parameters in the time and frequency domain features that can obviously represent the healthy operation state of the bearing are selected to construct degradation features, and the maximum-minimum normalization method is used for eigenvalue processing.

(2) Deep feature enhancement: The normalized feature signal is used as the input of DNN, and use its self-learning ability to extract and enhance the deep feature of the input signal, mainly including two stages of pre-training and model optimization. Firstly, the network parameters are initialized by pre-training, then the parameters are optimized by back propagation and Adam optimization function, and Dropout technology is applied to prevent over-fitting, so as to obtain the optimal network model. Finally, the enhanced degradation features are extracted and divided into training sets and test sets.

(3) Regression prediction: Take the characteristics of the training set as the input of GDBT, set relevant parameters, carry out iterative operation, and output the regression prediction value of the RUL of bearings.

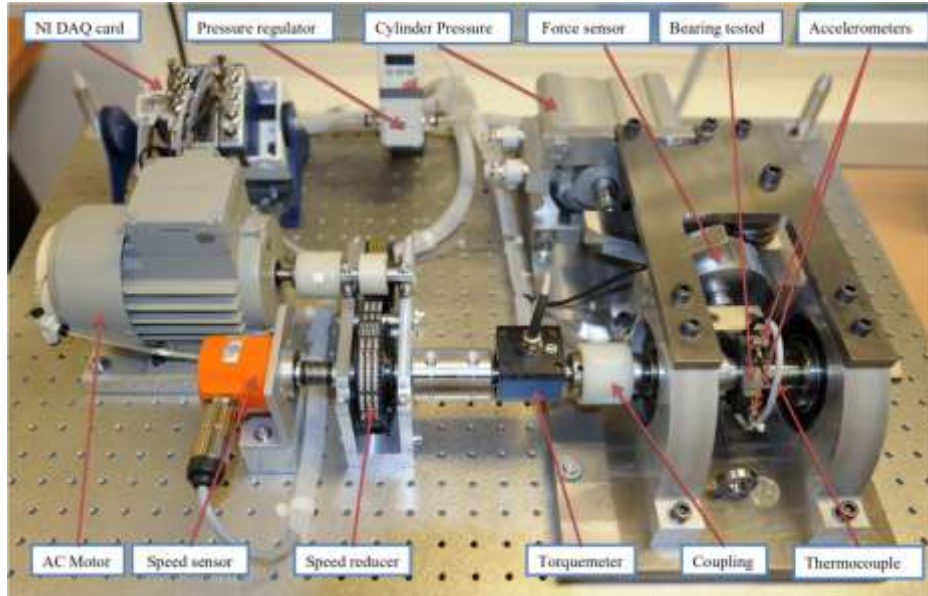
(4) Test set verification: The characteristics of the test set were input into the trained GDBT prediction model to calculate the corresponding bearing life prediction results.

## IV. Experimental Results and Analysis

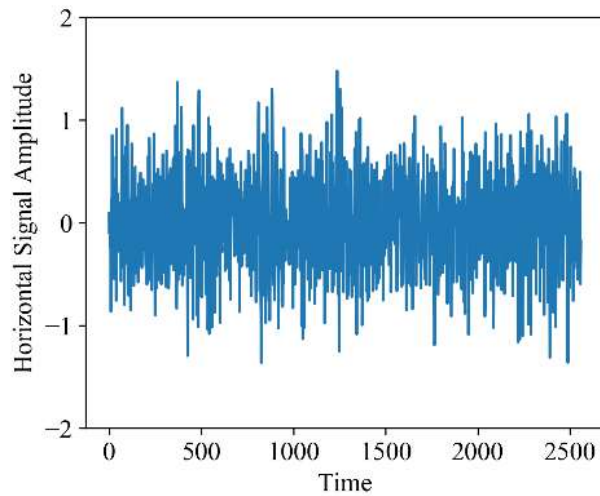
### 4.1 Experimental environment setting

The experimental data in this paper come from the data set of bearing life cycle provided by FEMTO-ST Institute, ISSN: 0010-8189  
© CONVERTER 2020  
www.converter-magazine.info

which is obtained by PRONOSTIA. The experimental platform is shown in Fig. 4. The horizontal and vertical vibration signals of one bearing are shown in Fig. 5 and 6. The experimental platform is Windows10/Visual Studio 2019/TensorFlow-GPU/Anaconda3, and the computer is configured as Intel i7-9750H CPU, 16.00 GB RAM, 64-bit operating system. A simulation program based on Python is designed, which is verified by 10-fold cross-validation method and compared with DNN, GBDT, SVR, RF, DT and other models.



*Fig 4: The PRONOSTIA experimental platform*



*Fig 5: Horizontal vibration signal of bearing*

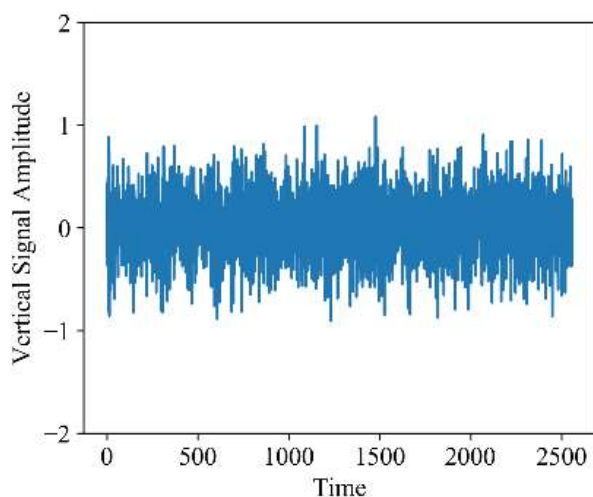


Fig 6: vertical vibration signal of bearing

#### 4.2 Experimental parameter setting

The structural design of deep neural network has a great influence on its performance. More hidden layers and hidden units can improve the prediction accuracy, but at the same time, it will increase the model complexity and the amount of calculation. To improve the correlation of the above 18-dimensional degradation features, this paper constructs a fully connected DNN with bearing degradation features as input data. All 10 hidden layers are connected, and the number of neurons is set to 500, 300, 200, 150, 100, 80, 50, 30, 18 and 1 respectively in hidden layers. The output was trained for 18000 times to generate DNN prior model features. The loss function use mean square error (MSE), which can represent the difference between the tag value and CNN network prediction value. Adam optimization function is used to update the weight and offset. At the same time, the learning rate attenuation mechanism is adopted to conduct iterative training, in which the minimum learning rate  $lr_{\min}$  and the maximum learning rate  $lr_{\max}$  are respectively  $lr_{\min} = 0.001$  and  $lr_{\max} = 0.01$ , and the attenuation rate is  $decay = 5000$ . The relationship between learning rate and convergence rate and loss value of DNN model is shown in Fig. 7. Each CNN model was trained for 10,000 times. During the training process, the mean absolute error (MAE) and root mean square error (RMSE) are used to evaluate the performance of the training model and preserve the CNN training model with the best performance.

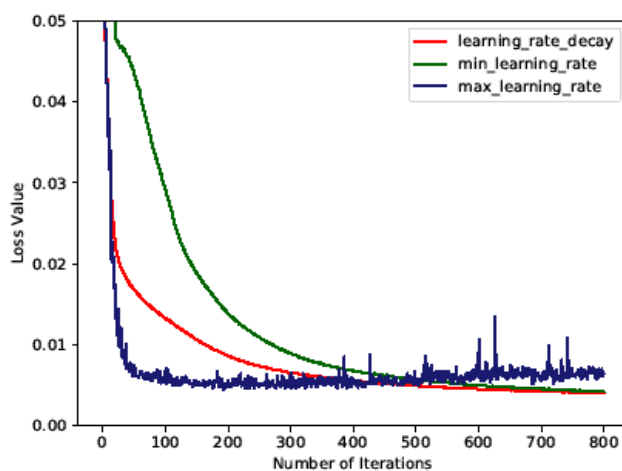


Fig 7: Relationship between learning rate and convergence rate and loss value of DNN model



The prediction performance evaluation indexes to evaluate the performance of the design method are shown in Eq. (12)-(15).

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (12)$$

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (13)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (14)$$

$$R\text{-square} = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (\bar{y}_i - y_i)^2} \quad (15)$$

Where,  $\hat{y}_i$  represents the real lifetime of sample point  $i$ ;  $y_i$  means the predicted life value of the sample point  $i$ ;  $i=1,2,\dots,N$ , is different sample points;  $N$  represents the predicted data points number.

### 4.3 Results and analysis

For the test data set, the collaborative prediction method of multi-bearing RUL based on DNN and GBDT regression designed above was adopted to predict the RUL of four bearings. The results of bearing RUL prediction are shown in Fig. 8.

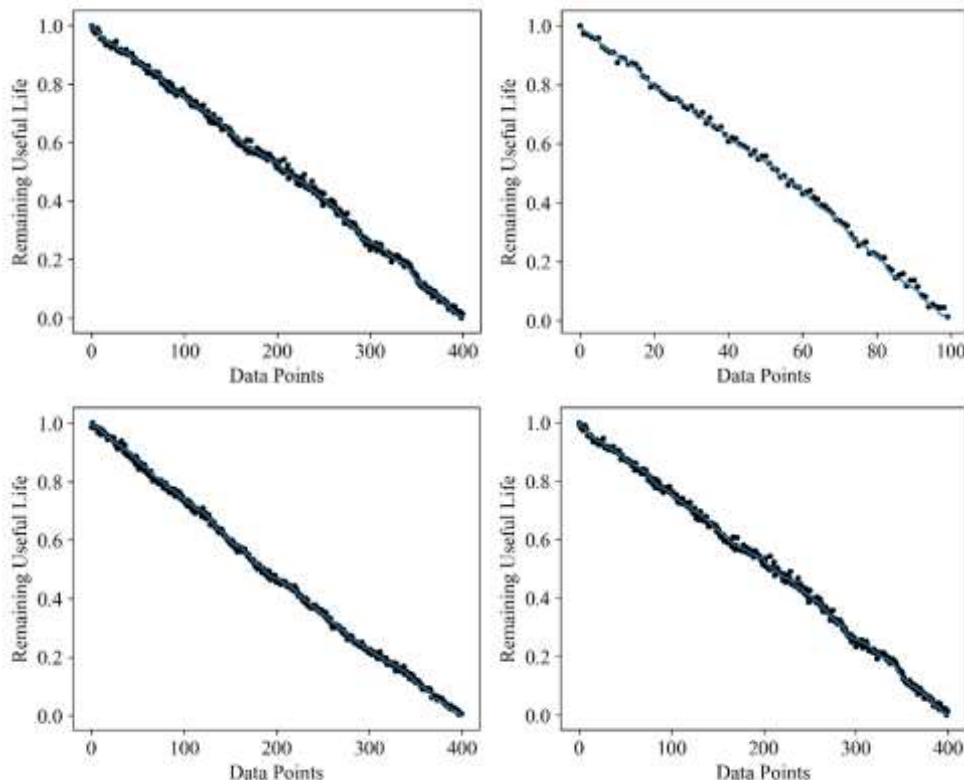


Fig 8: Prediction result of bearing RUL

The abscissa represents the sampled bearing data points, the ordinate represents the standardized remaining life of the bearing, the solid line is the real life value of the bearing, and the point is the RUL predictive value. From the

prediction results of the bearing RUL in Fig. 8, it can be obtained that the prediction value of this method has a good fitting effect with the real service life value.

In order to quantitatively analyze the above results, RMSE, MAE, MSE and R-square are selected as evaluation indicators to evaluate the method in this paper, and compared with DNN, GBDT, SVR, RF, and DT methods. The ratio of evaluation is shown in Table 1.

Table 1 Evaluation contrast ratio of prediction results

Contrast method	Evaluation index			
	RMSE	MAE	MSE	R-Squared
Method in this paper	<b>0.0469</b>	<b>0.0206</b>	<b>0.0022</b>	97.2646%
DNN	0.0584	0.0386	0.0053	<b>99.7938%</b>
GBDT	0.0762	0.0344	0.0058	92.0178%
SVR	0.0989	0.0549	0.0098	85.7876%
RF	0.0751	0.0290	0.0056	92.0222%
DT	0.1019	0.0284	0.0103	86.6576%

According to the evaluation pair ratio of the prediction results of the six methods in Table 1, it shows that the RMSE, MAE and MSE values of the prediction results calculated by the proposed method are all lower than those of DNN, GBDT, SVR, RF and DT methods, indicating that the collaborative prediction method based on DNN and GBDT has smaller prediction error and higher prediction accuracy. Compared with GBDT, SVR, RF and DT, the R-square of the prediction results of the collaborative prediction method is closer to 1, second only to DNN method, which shows that this method has good fitting effect.

Therefore, the collaborative prediction method of bearing RUL based on DNN and GBDT regression proposed in this paper has high prediction accuracy and calculation efficiency, and is expected to be applied to fault prediction and evaluation of electromechanical equipment.

## V. Conclusion

A collaborative prediction method of bearing RUL based on DNN and GBDT regression is proposed in this paper. To overcome the influence of the numerical range of different characteristic parameters, reduce the data dimension and improve the efficiency of the algorithm, the parameters that can obviously represent the healthy running state of the bearing in the vibration signal are selected as degradation characteristics, and the 18 dimensional degradation characteristic index based on the bearing vibration signal is constructed. Based on DNN, a prior feature model is generated and used for GBDT regression prediction. The results indicates that the prediction method has advantages of higher precision and fitting effect. It is expected to be applied to health failure prediction and health management of electromechanical equipment.

## Acknowledgements

This research was supported by Key Scientific and Technological Project of Henan Province (192102210222), Key Scientific Research Projects of the Higher Education Institutions of Henan Province (19B460003), and Doctoral Fund of Henan Institute of Technology (KY1750).

## References

- [1] F. Wang, J.K. Gu, R. Wang, "Bayesian-extension diagnosis model of equipment fault in ship integrated power system," *Journal of Ordnance Equipment Engineering*, vol. 40, no. 02, pp. 189-193, 2019.
- [2] H. Ahmed, M. Wong, A.K. Nandi, "Intelligent condition monitoring method for bearing faults from

- highly compressed measurements using sparse over-complete features,” *Mechanical Systems & Signal Processing*, vol. 99, no. 15, pp. 459-477, 2018.
- [3] L. Ren, J. Cui, Y. Sun, et al., “Multi-bearing Remaining Useful Life Collaborative Prediction: A Deep Learning Approach,” *Journal of Manufacturing Systems*, vol. 43, no. 02, pp. 249-256, 2017.
- [4] B. Wang, Y. Lei, N. Li, et al., “A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings,” *IEEE Transactions on Reliability*, vol. 69, no. 01, pp. 401-412, 2018.
- [5] W. Wei, J. Hu, J. Zhang, “Prognostics of Machine Health Condition using an Improved ARIMA-based Prediction method,” *2nd IEEE Conference on Industrial Electronics and Applications*, IEEE Press, 2007, pp. 1062-1067.
- [6] K. Liu, N.Z. Gebraeel, J. Shi, “A Data-Level Fusion Model for Developing Composite Health Indices for Degradation Modeling and Prognostic Analysis,” *IEEE Transactions on Automation Science & Engineering*, vol. 10, no. 03, pp. 652-664, 2013.
- [7] D. Wu, C. Jennings, J. Terpenney, “A comparative study on machine learning algorithms for smart manufacturing: Tool wear prediction using random forests,” *Journal of Manufacturing Science and Engineering*, vol. 139, no. 07, pp. 1-9, 2017.
- [8] M. Ma, c. Sun, x. Chen, “Discriminative Deep Belief Networks with Ant Colony Optimization for Health Status Assessment of Machine,” *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 12, pp. 3115-3125, 2017.
- [9] J.D. Zheng, J.S. Cheng, Y. Yang, et al., “A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination,” *Mechanism and Machine Theory*, vol. 78, no. 16, pp. 187-200, 2014.
- [10] R.M. Hasani, G. Wang, R. Grosu, “An automated auto-encoder correlation-based health monitoring and prognostic method for machine bearing,” *ArXiv*, no. 03, pp. 187-200, 2017.
- [11] P. Wang, R.X. Gao, R. Yan, “A deep learning-based approach to material removal rate prediction in polishing,” *CIRP Annals*, vol. 66, no. 01, pp. 429-432, 2017.
- [12] L. Liao, W. Jin, R. Pavel, “Enhanced restricted Boltzmann machine with prognosticity regulation for prognosticism and health assessment,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7076-7083, 2016.
- [13] C.L. Zhang, S.F. Ding, J. Zhang, et al., “Parallel stochastic configuration networks for large-scale data regression,” *Applied Soft Computing*, vol. 103, no. 05, pp. 1-10, 2021.
- [14] L. Liu, Y.J. Liu, S.C. Tong, “Neural Networks-Based Adaptive Finite-Time Fault-Tolerant Control for a Class of Strict-Feedback Switched Nonlinear Systems,” *IEEE Transactions on Cybernetics*, vol. 49, no. 07, pp. 2536-2545, 2018.