Image Feature-Based for Bearing Health Monitoring with Deep-Learning Method

Pengfei Shan*, Peng Hou, Honghong Ge, Linming Yu, Yang Li and Le Gu

Intelligent Equipment Research Laboratory
HRG (shandong) Intelligent Equipment Research Institute
Jinan, China
Email: shanpf@hitrobotgroup.com

***Abstract*—In the real industrial application, the problem of bearing health condition recognition and remaining useful life estimation is still confronted with many challenges. In this paper, a image feature-based combined with deep learning method is proposed for bearing online health monitoring. The primary difficult issue involved in this task is to automatic recognize the health condition and fault type and estimate the remaining useful life with high accuracy and feasibility. In order to find out the early fault of the bearing, the acceleration envelope detection algorithm is carried out which transform the time domain signal to frequency domain. Based on the envelope spectrum, the convolutional neural network (CNN) is adopted to recognize the health condition of bearing. The energy of the envelope spectrum is used as the health indicator and the remaining useful life is estimated by the long short time memory network (LSTM). The effectiveness of the proposed method is verified by the XJTU-SY datasets [1].**

Keywords: envelope spectrum; feature learning; health monitoring; remaining useful life; CNN; LSTM

I. INTRODUCTION

In recent years, with the development of the high-end equipment and the progress of computer technology, the predictive maintenance technique has raised the attention of many researchers. Rolling bearing is one of the most significant components of the machine. Due to the severe working environment, the bearing is prone to failure. If we can not conduct the replacement of the bearing with failure in time, it will make the machine shutdown which can bring out the economy losses. Therefore, the bearing health monitoring becomes a concerned issue of many manufacturers. With the rapid development of data collection and analysis technology, the data-driven method in bearing health monitoring has gained a lot of achievements. In recent years, deep learning has become a rapidly growing research direction in predictive maintenance applications [2]. The bearing health monitoring can include the health condition recognition, fault type recognition and the remaining useful life estimation. Traditional bearing fault diagnosis is mainly conducted by the following procedures. Firstly, the time domain features, frequency domain features and time-frequency features are extracted from the raw vibration data, and the features are selected by the feature dimension reduction method. The selected features are confused by the machine learning algorithm for health condition assessment and fault diagnosis. [3] used the designed features and support vector machine (SVM) for bearing fault recognition.[4] used artificial neural networks (ANN) based on empirical mode decomposition (EMD) for bearing fault classification. [5] used the raw time series data as the input of the CNN model to classify the fault type and achieved a good result. [6] used the fast Fourier transform (FFT) spectrum as the input of the recurrent neural networks to diagnose the fault type of the bearing. However, the traditional method needs manually design features and select primary features which rely on the expert knowledge and cost the labor time. The method from raw data input into the network will cost the running time especially when the sample contains a great volume of data. The method which used the FFT spectrum as the input of the network can not recognize the early fault of the bearing, especially when it is used in the condition to recognize the bearing condition whether normal or early failure. Traditional remaining useful life estimation method contains the following procedures. First of all, the health indicator is constructed based on the health assessment algorithm which fused the different selected features. Secondly, the time series prediction algorithm is adopted to predict the future trend of the health assessment curve a few steps later. Finally, the remaining useful life is estimated by calculating the time between the current time and the predictive failure time. This method not only relies on the selected features, data fusion algorithm but also depends on the long-time forecast ability.

Considering the problems mentioned above, in this paper, the envelope spectrum feature is learned by the CNN model and the output of the network is regarded as the health condition index. Once the proposed CNN model is trained, the model can be deployed for bearing health condition recognition. Envelope spectrum analysis is recognized as one of most reliable and feasible techniques in industrial bearing diagnosis application [7]. From the envelope spectrum we can find out the early fault drowned in the high frequency signal. In the remaining useful life estimation part, the acceleration envelope energy is used as the health indicator, the long short time memory networks (LSTM) is used for energy trend prediction.

The rest of this paper is organized as follows. Section II presents the theoretical background of the related work. In section III, the proposed health monitoring method is addressed. In section IV, the model architecture is described and the

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

details of the verification experiment is addressed. The paper ends with including remarks in Section V.

II. THEORETICAL BACKGROUND

A. Acceleration Envelope Detection Algorithm

Acceleration envelope spectrum can reveal the fault in bearing earlier stage of development. The early stage fault signal of bearing is often manifested in high frequency part. To solve this problem, the high frequency vibration signal component is filtered out by the band-pass filter. After filtering the signal, the Hilbert transform is used to extract the envelope signal which is further analyzed in the frequency domain. The flowchart of the algorithm is described as Fig.1.

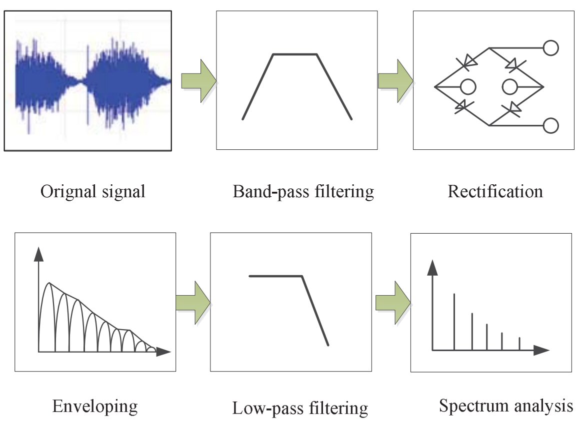

Fig.1. The flowchart of the envelope detection algorithm

Hilbert transform is usually used to obtain analytic signals [8]. According to this principle, the Hilbert transform can be applied to demodulate the narrowband signals. The Hilbert transform envelope demodulation algorithm is illustrated as follows.

For the vibration signal $x(t)$, after the Hilbert transform equation (1) can be obtained.

$$x(t)'=\int_{-\infty}^{+\infty}\frac{x(t)}{t-\tau}d\tau \tag{1}$$

Then the analytic signal can be constructed as Eq. (2).

$$z(t)=x(t)+jx(t)' \tag{2}$$

The conjugate analytic signal of $z(t)$ can be expressed as Eq. (3).

$$z(t)'=x(t)-jx(t)' \tag{3}$$

By the quadrature operation between the analytic signal and the conjugate analytic signal, a new synthetic signal can be obtained, therefore, the magnitude of the analytic signal or conjugate analytic signal can be used as envelope signal.

$$env=\sqrt{z(t)z(t)'} \tag{4}$$

B. Convolutional Neural Network (CNN)

CNN[9] was firstly proposed by LeCun for image processing, which was expressing for their two prominent properties: spatially-shared weights and spatial pooling. In recent years, CNN has achieved a great success in computer vision field, such as the face recognition and object detection. However, there are few researches involved in industrial machine health monitoring application based on CNN. Generally speaking, if different bearings have the same speed and type parameters, they will show similar features on the spectrum when they have the same fault, although the amplitude of spectrum may be different. Besides, different faults will have different patterns expressed in their envelope spectrum. Therefore, the CNN could be a feasible model for image-feature learning and fault recognition.

The convolutional neural network (CNN) generally contains the convolutional layers, pooling layers and fully-connected layers. In the convolutional layers, assuming the number of the convolutional filter is k, each filters can extract on type of features, so the output of the extracted features after the convolutional operation is k.The convolutional operation can be expressed by:

$$Y_k=f(W_k*X+b) \tag{5}$$

Where W_k is the k th convolutional kernel with the size $k_1\times k_2$, X is the input data with the size $m\times n$, b is the bias. '*' denote the convolutional operation. If the slide stride is 1×1 and no padding operation the size of output for each filter will be $(m-k_1+1)\times(n-k_2+1)$, f denotes the activation function, the commonly used function can be 'ReLu'.

In the pooling layers, the k th pooling filter size is $p_1\times p_2$, if the stride is 1×1， and no padding operation then the size of the output for each filter after pooling operation will be

$$(m-k_1+1-p_1+1)\times(n-k_2+1-p_2+1).$$

In the fully connected layers, the feature will be deeply fusion and classified after the regression and 'softmax' operation.

C. Long Short Term Memory Network (LSTM)

Long short term memory network has been proved to be the most stable and powerful model to capture the long range temporal dependencies in practical applications [10]. The core idea of the LSTM lies in the introduction of a few gates in each time step to control the past information flowed to the next time step. In each time step, the hidden state $a^{<t>}$ can be updated by the current time series $x^{\{t\}}$,previous hidden state $a^{<t-1>}$, the update gate $\Gamma_u^{<t>}$, the forget gate $\Gamma_f^{<t>}$, the output gate $\Gamma_o^{<t>}$ and the memory cell $c^{<t>}$. The relationship among the can be given as follows:

$$\Gamma_f^{<t>}=\sigma(W_f[a^{<t-1>},x^{<t>}]+b_f) \tag{6}$$

$$\Gamma_u^{<t>}=\sigma(W_u[a^{<t-1>},x^{<t>}]+b_u) \tag{7}$$

$$\tilde{c}^{<t>}=\tanh(W_c[a^{<t-1>},x^{<t>}]+b_c) \tag{8}$$

$$c^{<t>} = \Gamma_f^{<t>} \circ c^{<t-1>} + \Gamma_u^{<t>} \circ \tilde{c}^{<t>} \tag{9}$$

$$\Gamma_o^{<t>} = \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o) \tag{10}$$

$$a^{<t>} = \Gamma_o^{<t>} \circ \tanh(c^{<t>}) \tag{11}$$

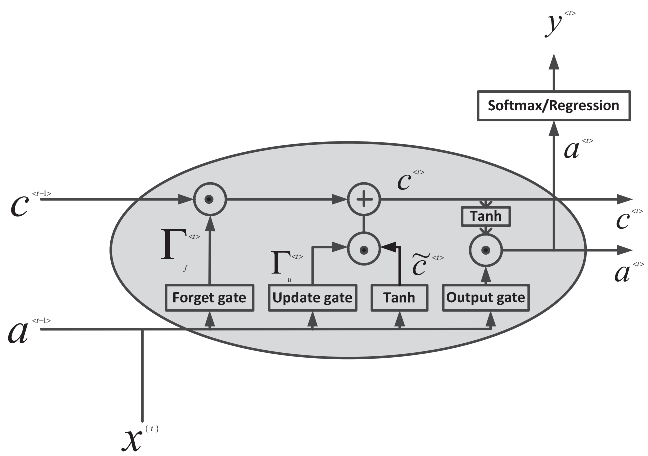

Fig. 2. Lstm cell structure

The predicted value $y^{<t>}$ can be obtained by linear regression or softmax function, and the networks can be trained by the back propagation through time (BPTT) algorithm. Once the training process is completed the model can be applied in time series prediction.

III. Proposed Bearing Health Monitoring Method

In this paper, the acceleration envelope detection algorithm is developed to obtain the envelope spectrum. Based on the envelope spectrum image, the condition recognition algorithm is developed. Based on the envelope energy the remaining useful life estimation algorithm is carried out. The proposed method mainly contains the following procedures, and the flowchart of the proposed method is demonstrated as Fig.3.

Step1: Obtain the time series data from the vibration sensor.

Step2: Transform the time series data into the envelope spectrum for each sample data.

Step3: Divide the data into the training samples and test samples.

Step4: Train the convolutional neural network using training samples

Step5: The bearing life time sample is input in the trained model.

Step6: With the time step goes on the bearing condition will progress from normal to failure, once the continuous failure is detected, the LSTM prediction model is adopted for remaining useful life estimation.

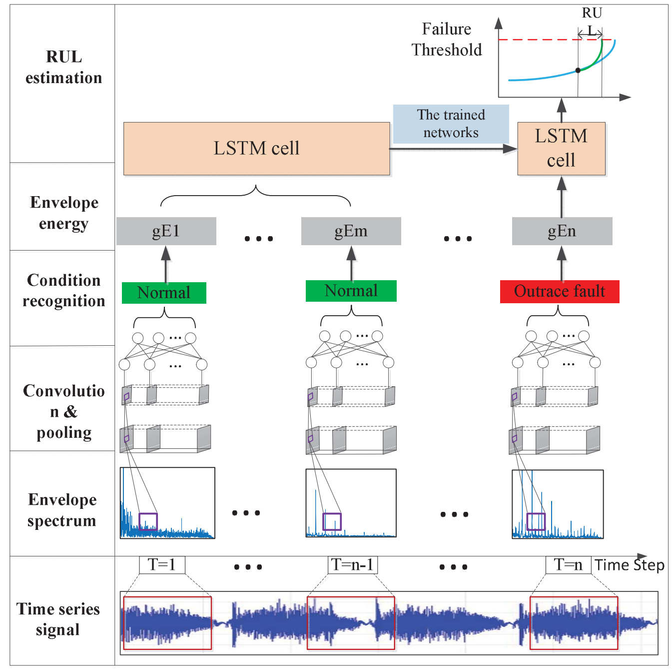

Fig. 3. The flowchart of the proposed method

IV. Experiment and Result

In this section, we will show a comparison experiment of the proposed condition recognition method with several traditional methods and the accuracy of these algorithms will be addressed. Besides, the remaining useful life will be estimated by the long short term memory networks.

A. Experiment Setup and Dataset Description

In the experiment, 15 bearings were tested under different operating conditions (i.e.,35Hz and 12KN, 37.5Hz and 11KN, 40Hz and 10KN.) and the fault element contains the outer race fault, inner race fault and cage fault [1]. In the experiment, the bearing2-2, bearing2-3 and bearing3-4 are considered to be the dataset used for the proposed method validation and the normal condition, the outer race fault condition, the inner race fault condition, the cage fault condition are included in the dataset.

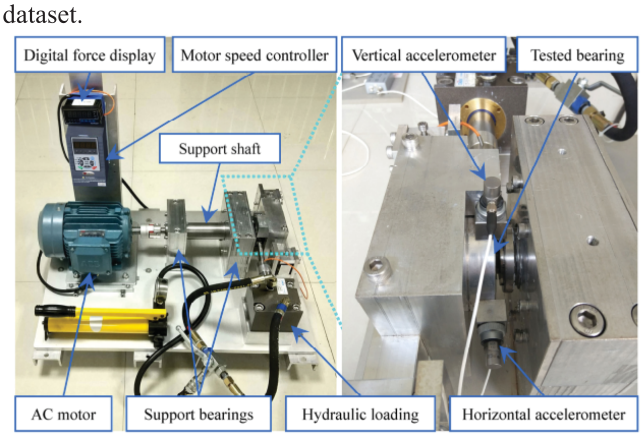

Fig.4. Tested of rolling element of bearings

B. The CNN Model and Experiment Result

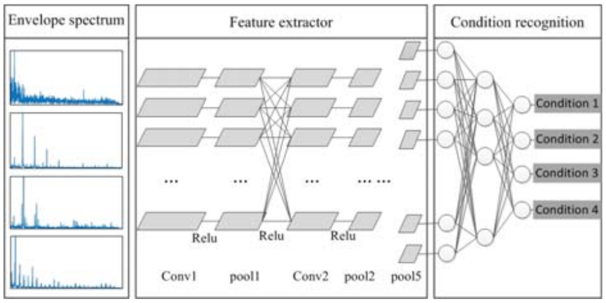

Fig.5. The proposed CNN model

In order to obtain the envelope spectrum, the envelope detection algorithm is conducted. In the envelope spectrum image, the x axis is constrained in [0,1000] Hz, the y axis restricted between 0 and the maximum amplitude of the envelope spectrum in the frequency [0,1000] Hz. In the experiment, the size of the picture is set to be 32×64.

As shown in Fig.5, the proposed CNN model contains five layers, and each one contains convolutional layers, rectified liner unit (ReLu), maxpooling layers for feature extraction and two fully connected layers for condition recognition. For our proposed CNN model five convolutional layers are designed whose filter number are set to 64,128, 256,512 and 1024, kernel size are all 2×2 ,strides are all 1×1, and padding are all configured as 'same'. For the five pooling layers, their pool sizes are all 2×2 and strides are all 2×2. After each convolution and pooling operation the BatchNormalization will introduced to avoid overfitting.

As shown in TABLE I, four conditions of the dataset are described. And each condition data contains 80 samples, which 60 samples are used to the model training and the rest 20 samples for the model testing. As shown in TABLE II, a comparative experiment is designed for the proposed method validation. The first three models are designed for different parameters, the fourth model is carried out with the traditional method which is implemented from feature extraction to classification.

TABLE I THE EXPERIMENT DATASET

Condition	Train.No	Test.No	Total.No
Normal	60	20	80
Outer race fault	60	20	80
Cage fault	60	20	80
Inner race fault	60	20	80

TABLE II THE EXPERIMENT RESULT

Model	Parameters	Accuracy
CNNs	Batch_size=30, lr=0.01, epochs=10	83%
CNNs	Batch_size=30, lr=0.001, epochs=15	95%
CNNs	Batch_size=30, lr=0.0001, epochs=15	**100%**
SVM	Kernel=rbf	60%

C. The LSTM Model and Experiment Result

In the remaining useful life estimation part, the stacked LSTM model will be addressed for energy trend prediction. Due to the bearing condition in the early degradation is relatively stable and in order to ensure the prediction accuracy, the RUL estimation will be implemented when the bearing condition is recognized as abnormal by the CNN model. The proposed LSTM model contains three LSTM cells and two fully connected layers. The hidden unit of the three LSTM cells are 100,50,50 respectively and the hidden unit of the two fully connected layers are 50 and 1 respectively. In the experiment the energy trend is predicted by two kind of ways. One method is the short term prediction which predict the next step value based on the current value and the trained networks and the other one is the long term prediction which predict the next 25 steps based on the current value and the trained networks. In the experiment, the bearing1_1 is used for the test dataset, as the Fig.8 shows , 80% of the whole life time data is used for training, the black line indicates the failure threshold, the red curve is the predicted value curve which obtained by predicting the next 25 steps from the prediction starting point, the green curve is the predicted value curve which means the one step method. The blue curve is the real spectrum energy trend. In the same condition, there are five groups of datasets and each dataset will have its own failure threshold. In this paper, the threshold is set by the following rules. Firstly, the bearing spectrum energy of five groups at the failure time will be calculated. Let the minimum value as M ,and then each spectrum energy can be calculated by:

$$value\_new = \frac{value}{M} \tag{12}$$

where the value means the spectrum energy value, and the value_new represents the normalized value. Therefore, in the same condition, the failure threshold will be greater than or equal to one. In the paper, the failure value is set to one.

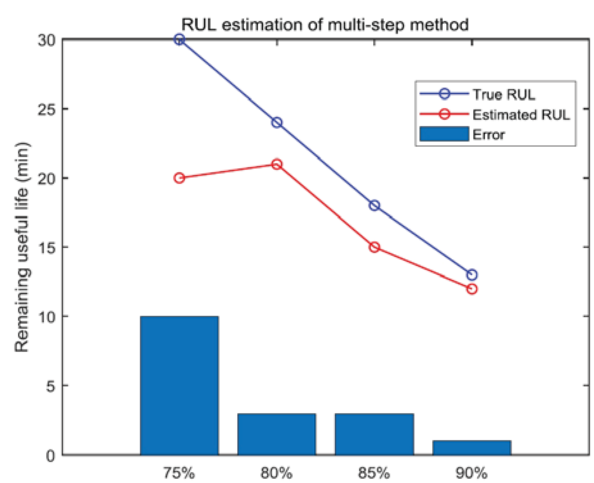

Fig. 6. RUL estimation of multi-step method

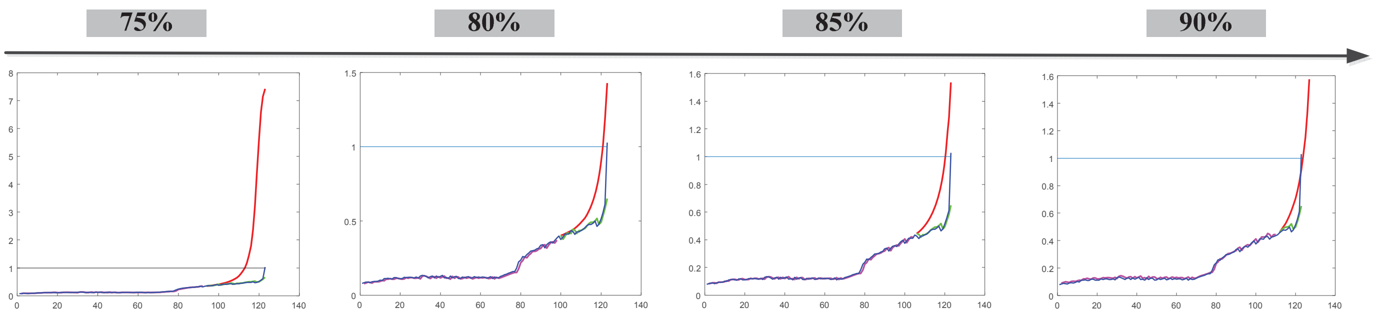

Fig.7. RUL estimation of different life stage

General speaking, the large range fluctuations of the energy curve will infulence the remaining useful life estimation result. Therefore, before implementing the remaining useful life estimation, the data smoothing and curve fitting technique to remove the small random fluctuations has been utilized. In this paper, the energy curve obtained by the envelope algorithm are smoothed by the locally weighted regression (LOESS) [11]. In LOESS, every smoothed data is calculated by the neighboring data with a given span. The weight in every data in the span is calculated by Eq. (13).

$$w_i = (1 - \left|\frac{x - x_i}{d(x)}\right|)^3 \tag{13}$$

where x is the predictor value associated with the value to be smoothened, x_i denotes the nearest neighbors of x in the span, and $d(x)$ represents the horizontal distance between x and the most distant predictor value in the span. LOESS then uses locally quadratic regression as a weighted linear least squares regression, determining the smoothed value through weighted regression at the predictor value.

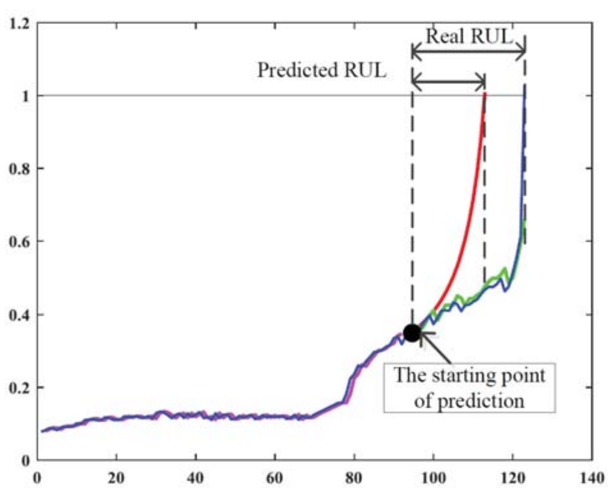

Fig.8. RUL estimation (training data: 80% of whole data)

TABLE III COMPARASIONS OF ONE STEP AND MULTI-STEP RESULT

Error	75%	80%	85%	90%
RMSE1	13.2945	1.0581	1.2763	0.4548
RMSE2	0.3810	0.3873	0.3871	0.3837

TABLE IV RUL ESTIMATION OF MULTI-STEP PREDICTION METHOD

RUL and Error	75%	80%	85%	90%
Real RUL	30min	24min	18min	13min
Estimated RUL	20min	21min	15min	12min
error	10min	3min	3min	1min

In TABLE III the comparative result between one step and multiple steps prediction method is conducted. In the table, the train data contain four groups, i.e. 75% ,80%, 85%, 90% of the whole life time data respectively. To quantitatively evaluate the performance of the one step and many steps prediction method, root mean squared error (RMSE) is adopted to measure the prediction precision. RMSE1 represents the many step prediction method, RMSE2 represents the one step by one step prediction method. RMSE can be calculated by the following equation.

$$RMSE = \sqrt{\frac{1}{n}(y_i - \overline{y_i})^2} \tag{14}$$

where y_i represents the real target value, and the $\overline{y_i}$ represents the predicted target value, n represents the number of samples.

As demonstrated in TABLE IV , the remaining useful life of multiple step prediction method experiment is carried out. As shown in the table, with increase of the training data percentage the remaining useful life estimation error will become smaller.

V. Conclusion

In this paper, an image feature-based method for bearing health monitoring is proposed. The proposed method mainly implemented by the acceleration envelope spectrum combined with deep learning method. It mainly tackles the bearing health monitoring problem, from condition recognition and remaining useful life estimation automatically. In this paper, the acceleration envelope detection algorithm is developed. Then, the envelope spectrum is obtained by the developed algorithm.

After that, the convolutional neural networks were adopted for the bearing condition recognition. Additionally, the spectrum energy will be calculated based the result of envelope detect algorithm. Finally, the remaining useful life is estimated by the long short term networks and the validation experiment is carried out and verified the effectiveness and feasibility of the proposed method.

However, there are still some limitations in the proposed method, In this study, the failure threshold is set by five groups with the same working conditions, as a matter of fact, this threshold may be more accuracy with a larger number of data, Therefore, in the future, the bearing threshold will be defined more systematically with a great number of whole life-time data.

REFERENCES

[1] http://gr.xjtu.edu.cn/web/yaguolei

[2] Rui Z, Ruqiang Y, Zhenghua C, et al. Deep learning and its applications to machine health monitoring[J]. Mechanical Systems and Signal Processing, 2019, 115:213-237.

[3] Zheng H, Zhou L. Rolling element bearing fault diagnosis based on support vector machine[C]// International Conference on Consumer Electronics. IEEE, 2012.

[4] Dubey R, Agrawal D. Bearing fault classification using ANN-based Hilbert footprint analysis[J]. Science Measurement And Technology Iet, 2015, 9(8):1016-1022.

[5] Sadoughi, Mohammakazem; Downey, Austin; Bunge, Garrett; Ranawat, Aditya; Hu, Chao; and Laflamme, Simon, "A Deep Learning-based Approach for Fault Diagnosis of Roller Element Bearings" (2018). Civil, Construction and Environmental Engineering.

[6] Jiang H, Li X, Shao H, et al. Intelligent fault diagnosis of rolling bearing using improved deep recurrent neural network[J]. Measurement Science and Technology, 2018

[7] Li Z T, Li H. EMD and Envelope Spectrum Based Bearing Fault Detection[J]. Advanced Materials Research, 2012, 459:5

[8] Feldman M. Hilbert Transform Applications in Mechanical Vibration[J]. John Wiley,Sons, 2011.

[9] Le Cun Y. Handwritten digit recognition with a back-propagation network[J]. Neural Information Processing Systems, 1990, 2.

[10] Qiao, Huihui and Wang, Taiyong and Wang, Peng, et al. A TimeDistributed Spatiotemporal Feature Learning Method for Machine Health Monitoring with Multi-Sensor Time Series[J]. Sensors, 2018

[11] Cleveland W S, Devlin S J. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting[J]. Journal of the American Statistical Association, 1988, 83(403):596-610.