Stacked Convolutional LSTM Models for Prognosis of Bearing Performance Degradation

Wang Du
Institute of Electrical and Mechanical Engineering
Guilin University of Electronic Science and Technology
Guilin, China
duwang1631@163.com

Yanxue Wang
Institute of Mechanical-Electronic and Vehicle Engineering
Beijing University of Civil Engineering and Architecture
Beijing, China
yan.xue.wang@gmail.com

***Abstract*—Bearing fault is one of leading causes among industry machine failures. Detection and diagnosis of bearing faults is therefore pretty significant for health monitoring of mechanical equipment. The latest years has witnessed more and more attention to deep learning techniques for prediction and health management (PHM) of bearings. The degradation performance induced by vibration features can be predicted with less requirements on prior knowledge. However, many existing bearing RUL prediction methods have the following two shortcomings: (i) the effects of the entire bearings running time and continuous variation are often ignored; (ii) gradient vanishing or gradient explosion is often involved in learning long term dependency in traditional neural networks. To overcome these problems, a deep structure of bearing degradation prognosis based on staked convolutional and long-short-term memory (CNN-LSTM) has been proposed in the paper. Firstly, convolution neural networks are applied to extract features from the original vibration signals, and then the extracted features are fed to the stacked long-short-term memory networks through the connection layer. Finally, the bearing performance degradation trend is estimated. In this experiment, the bearing dataset of NASA is utilized to verify the prognosis performance of the proposed method. The experiment results show the prognosis effectiveness of the proposed method is prior to traditional CNNs and plain LSTM networks.**

Keywords-bearing performance degradation prognosis; stacked convolutional LSTM models; convolutional neural networks; long-short-term memory networks

I. INTRODUCTION

Rolling bearings are among most essential parts in industry machine, which are widely applied in modern industry [1]. Therefore, reliable fault diagnosis and prognosis schemes of bearings are very significant to guarantee the safety of system operation and human operators, and to minimize the risk of irreversible damage to components [2].

In the past decades, significant advances in theoretical and applied research have occurred in the area of fault diagnosis. Currently, there are mainly three paradigms for fault diagnosis of bearing performance degradation: data-driven methods [3], model-based methods, and hybrid methods. model-based methods minimize the need for a priori data and can perform online, but they require accurate mathematical model of the system. In contrast, the data-based method does not need much understanding of the inherent fault mechanism of the system., which is mainly based on the dataset collected by sensors. Therefore, data-driven methods can be very useful when it is not available to obtain a mathematical model of linear or non-linear systems, which has led to more and more attention in recent years [4, 5].

Currently, people have developed many data-driven methods combined with new information technology [6]. These methods usually utilize machine learning and pattern recognition techniques to build prediction models based on historical data to predict future values [7, 8]. However, many existing bearing RUL prediction methods have the following two shortcomings: (I) the effects of the entire bearing running time and continuous variation are often ignored; (II) gradient vanishing or gradient explosion is often involved in learning long term dependency in traditional neural networks [9]. To overcome these problems, a deep framework for bearing degradation trend prognosis based on staked convolutional and long-short-term memory networks (CNN-LSTMs) is applied in the paper.

The paper attempts to construct convolutional long-short-term memory (CNN-LSTMs) to prognosis bearing performance degradation. The outline of this article is as below: Part II describes essential conceptions of CNN, RNN, LSTM and CNN-LSTM briefly. Part III presents experimental results of experimental validation and method comparison. And Part IV presents the conclusion of the experiment.

II. ESSENTIAL CONCEPTIONS

The distinctive feature of machine learning is a multi-layer sensor with multiple hidden layers based on deep neural network (DNN) and deep learning can be regarded as a subset of machine learning .There exist several neural networks of

This research was funded by National Natural Science Foundation of China under Gant No. 51875032, 51475098 and 61463010) and Guangxi Natural Science Foundation (2016GXNSFFA380008).

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

deep learning, such as CNN,LSTM, convolutional LSTM and so on [10].

A. Convolutional Neural Networks

The input raw sensor signals can be fed to neural networks. And features obtained from the original sensor signals sending to the neural networks usually can enhance performance. However, finding enough features requires much expert knowledge, which will inevitably restrict an in-depth study of features [11]. In the purples of solving the problem, convolutional neural networks are proposed.

Convolutional neural networks (CNNs) were originally proposed for image processing by [12]. CNN, as a subset of deep learning, can utilize spatial shared weights and spatial subsampling (or pooling) as feature extractors, and apply stacked convolution operations to create multi-level and hierarchies of features step by step. Such a model has ability to learn features automatically at multiple levels (also known as "representation learning") [5].

The features of the raw signal can be extracted by a single-layer CNN, which is by a filter (or convolution kernel). The cell of CNN convolution network, is activated by the result of convolution operation of kernel and signal. By calculating the activation of a cell, the mode can be captured by the kernels. And during the supervised learning process, the kernel obtains optimization. Feature mapping is a series of cells (or layers) that share the same parameters (weight vectors and deviations). The activation results in kernel convolution throughout the input data [7].

The dimension of input data is able to decide the application of convolution. For a two-dimensional image time series, two-dimensional convolution kernels are often used for two-dimensional spatial convolution; and for one-dimensional time series (such as one-dimensional vibration signals), one-dimensional convolution kernel is often used for one-dimensional convolution [13]. In one-dimensional time domain, the convolution kernel acts as a feature detector to delete abnormal points and filter data.

In terms of mathematical formulas, the following is used to extract feature map by one-dimensional convolution operation:

$$a_j^{l+1}(\tau) = \sigma\left(b_j^l + \sum_{f=1}^{F^l} K_{jf}^l(\tau) * a_f^l(\tau)\right) = \sigma\left(b_j^l + \sum_{f=1}^{F^l}\left[\sum_{p=1}^{p^l} K_{jf}^l(p) a_f^l(\tau - p)\right]\right) \quad (1)$$

where $a_j^{l+1}(\tau)$ depicts feature mapping j of layer l . σ and F^l are a non-linear function and the number of feature maps separately. And b_j^l is a bias vector of layer l , K_{jf}^l denotes convolutional kernels of feature mapping f of l . And then, feature mapping j will be created in layer $l+1$. When processing sensor input data, the formula can be applied to each sensor channel of the data. independently.

The biggest difference between RNN and other deep learning methods such as DBN, SAE and CNN is that it considers the association between the samples context fully. The information of the sample context is represented with recurrent connection between the neural networks, which means the nodes between the hidden layers are connected. This connection takes into account the output of the current sample, not only affected by the current input information, but also influenced by historical information [10]. A typical RNN is shown in Fig.1. And this connection is expressed by the formula as below:

$$h_t = f(w_1 x_t + w_2 x_{t-1} + b_h) \quad (2)$$

where h_t denotes the hidden output at time t. f is an activation function. w_1 and w_2 are both weight matrix, the former is between input layer and hidden layer and the latter is between hidden layer at time t and at time $t-1$. b_h represents a bias. However, excessive recursion times is equivalent to an increase in the depth of the neural network, which makes the training time longer and causes gradients vanishing or exploding problem.

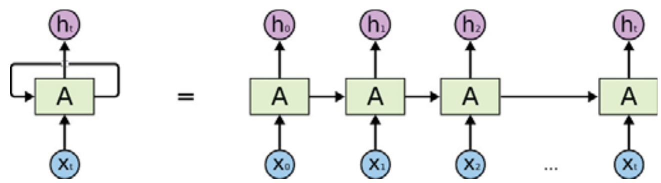

Figure 1. A recurrent neural network

Long-short-term memory neural networks (LSTM) are constructed to solve the problem , which is by designing several control gates to replace the hidden layer of general RNN [14].

B. Long-Short-Term Memory Netwoks

LSTM, as a variant of RNN, is close to time series. LSTM neural networks have several memory blocks connected recursively. Each and every block contains a connection with itself and three other memory units (i.e., output gate, input gate and forgetting gate)[14]. These gates work together to enable memory units of LSTM networks to receive and store long-term information, thus alleviating he gradual disappearance of gradients [15]. In various kinds of LSTM networks, a memory block diagram of a commonly used LSTM network is shown in Figure 2 [16].

As shown in Figure 2, the memory block contains several basic parts: one block input, one block output, one simple cell, three gates and several pipeline connections. Suppose N, M and x^t are the number of LSTM blocks, the number of inputs and the input vector at time t, respectively. The parameters of the model include all input weights W , recursive weights R , pipeline weights p and biases b.In addition, i^t , o^t , f^t and c^t are input gate, output gate, forget gate and memory cell separately. Suppose that the hidden state y_t on time step x_t is updated by the same time step t. And at the past time step, the hidden state is obtained from y_t. The vector function of a LSTM block can be defined as below [16]:

$$z^t = g(w_z x^t + R_z y^{t-1} + b_z) \quad (3)$$

$$i^t = \sigma(w_i x^t + R_i y^{t-1} + p_i \odot c^{t-1} + b_f) \quad (4)$$

$$f^t = \sigma(w_f x^t + R_f y^{t-1} + p_f \odot c^{t-1} + b_f) \quad (5)$$

$$c^t = z^t \odot i^t + c^{t-1} \odot f^t \tag{6}$$

$$o^t = \sigma(w_o x^t + R_o y^{t-1} + p_o \odot c^t + b_o) \tag{7}$$

$$y^t = h(c^t) \odot o^t \tag{8}$$

Where σ , g and h are individually pointwise nonlinear activation functions. Symbol $\odot$ represents point-by-point multiplication between two vectors. The sigmoid $\left(\sigma(x) = \frac{1}{1+e^{-x}}\right)$ is the block input and the hyperbolic tangent ($g(x) = h(x) = tanh(x)$) represents the activation of output. The output vector of the input gate in formula (4) is multiplied by the block input vector in formula (3) to get a cell, which can decide which sequence data is stored and which sequence data is forgotten in the current cell state of formula (6).

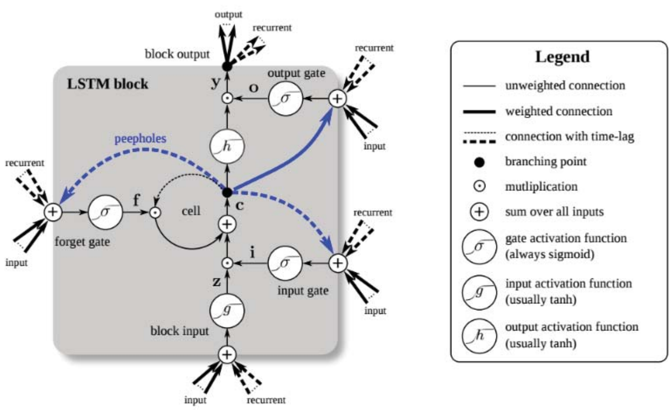

Figure 2. Diagram of LSTM memory block

C. Stacked Convulutional Long-Short-Term Memory Networks

Deep CNN and LSTM structures with multi-layer hidden layers can create higher-level features of time series data, and experimental conclusions also indicate rather good performance [9]. The layer-by-layer stacking mechanism can improve the function of the neural networks, and this method is also applied in the following.

CNN, as a feature extractor, can obtain the dimension of feature space from the input signal and extract the spatial feature vector, as shown in the earlier chapters. Thus, CNN is utilized to extract features in the ahead layers of the model. In addition, LSTM is suitable for obtaining time-domain features from time series data (e.g. one-dimensional vibration signals). Therefore, let LSTM layers be in the following layers. When the model calculates coming values, the final layer of the structure only requires the outputs from lower layers to calculate iteratively and then produce prognosis values. Therefore, for the sake of solving the problem of gradient disappearance or gradient explosion in the long-term information dependence of time series, LSTM layer is utilized in the last layers of the model.

A new deep networks structure, named superimposed convolutional LSTM neural network, is used for predicting future values in the paper. In the model, stacked CNNs extract the feature map from original vibration signals, and then the results will be fed up to stacked LSTM networks. Finally, the last LSTM layers present prediction values.

III. Experiments

The experimental performance of the proposed model (stacked convolution LSTM networks) is compared with that of pure CNN and LSTM networks separately.

A. Bearing Test Rig and Datasets

The experimental data used in the paper can be acquired by the website of NASA Prediction Database [17]. Figure 3 illustrates the bearing test bench and sensor device. As illustrated in the figure , four Rexnord ZA-2115 double-row bearings are mounted on the same shaft . The rotation speed of the shaft is stable at 2000 rpm. A radial load of 6000 lbs. is applied to the shaft, and the spring mechanism is also utilized. All the bearings are lubricated forced with lubricating oil, the flow rate and temperature of which are regulated by the oil circulation system [18].

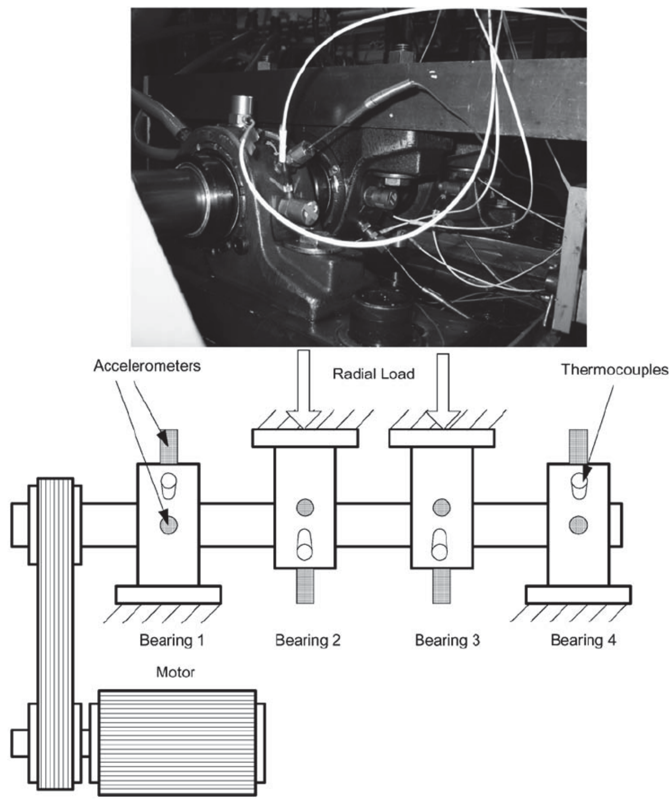

Figure 3. Bearing testbench and sensor placement

The experimental data is divided into two parts, namely training set and test set. The proposed model learns on the former, which contains known outputs. And the effectiveness of the proposed method is verified by the latter. Bearing failure can be characterized by root mean square (RMS) of vibration data, which is defined below:

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} \tag{9}$$

In the data specification, the data sets contain 984 files, each of which consisting of 20480 data points. And the dataset has a rate of 20.48kHz at 10-minute intervals. And original data points and its corresponding RMS are plotted in Fig.4 and Fig.5, separately. As shown in the Fig.4 and Fig.5, the curves are the RMS of the raw vibration signals and can be obtained through Eq. (9). Note that at the first 700 time series points, the amplitude is relatively stable. Then, an early failure arises in the 700th point. After the 700th point, the amplitude changes more and more dramatically, and reaches the maximum near the 975 time series point. At this time, it is the latest stage of the fault occurrence.

In this paper, for the sake of illustrating the robustness of the model, the prognosis performance of the model is tested in the stage of bearing smooth operation and the stage of fault deterioration, separately. For the former, we choose 1-300,300-600 data points as training data sets and 300-600,600-800 data points as validation data sets; for the latter, because of the drastic change trend in the worsening stage of bearing failure, we choose 600-800, 880-900, 900-920, 920-940 and 940-960 data points as training data sets; corresponding to this ,we select 800-900, 900-920, 920-940, 940-960 and 960-980 data points as validation data sets, respectively.

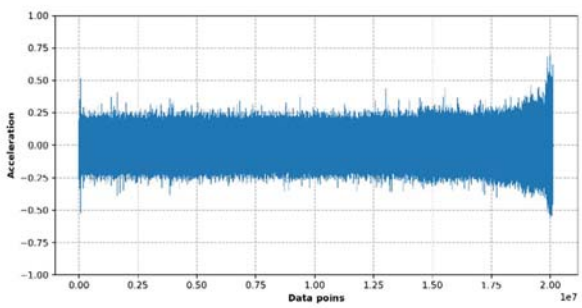

Figure 4. Primary Data Points

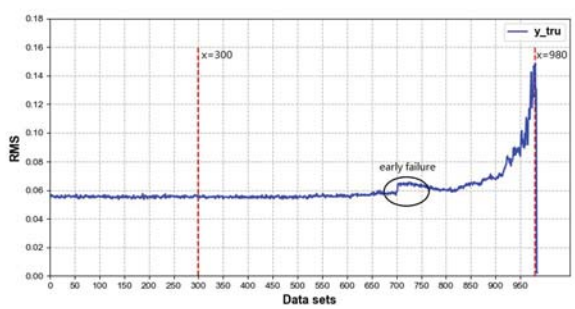

Figure 5. RMS of Primary Datasets

B. Evaluation Setup

To illustrate the validity of the proposed model in bearing data, a Keras 2.2.4 with a backend of TensorFlow 1.9.0. is used in the paper. We use the Python programming language to validate our experiments, which was on a note-book computer with an i7-8750 CPU of 2.2GHz and 16 GB main memory.

In this paper, we use two CNN layers and two LSTM layers. And the units of two CNN layer are 128 and 64, and both of LSTM layers are 64. And batch normalization method has been used to improve the prognosis effect of the proposed model. Moreover, the activation function of the output layer is set to a tanh function. In addition, the model is trained with Adam optimizer optimization algorithm. The training set is fed into the proposed model and the test set is used to detect the performance degradation trend of bearing predicted by the model. Both the original RMS and the predicted RMS of the data are illustrated in the following in Fig.6.

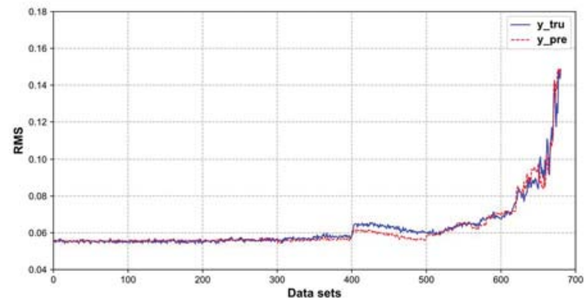

Figure 6. Prediction Result of Proposed method

For the sake of comparing the proposed method with other neural network methods in detail, two kinds of evaluation indexes are used in this paper. Both mean square error (MSE) and mean absolute error (MAE) are good indicators to measure prediction errors, which can mirror the actual state of prediction errors. The corresponding calculation formulas for these indicators are as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\overline{y_i} - y_i)^2 \tag{10}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\overline{y_i} - y_i| \tag{11}$$

where y_i manifests the real value and $\overline{y_i}$ denotes the predicted value.

The stacked convolutional LSTM model proposed in the paper, is compared with LSTM and CNN networks to illustrate the advantage of it in prognosing rolling bearing degradation performance. And the results have been shown in Table I, which illustrate and compare the losses of the proposed method, CNN and LSTM. Table I describes the loss value for predicting the smooth operation of bearings (data sets 300-600) and the deterioration of bearing faults (data sets 600-980).

According to the Table I, whether in the stage of bearing smooth operation or in the stage of fault deterioration, the prognosis effect of the stacked convolutional LSTM networks is much better than the others. Therefore, the experimental results show that the proposed method is effective in predicting the performance degradation trend of bearing faults.

TABLE I. LOSSES COMPARISION

	Proposed method		CNN		LSTM	
	MSE	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*
$e_{300-600}$	0.0017	0.0331	0.0019	0.0368	0.0123	0.0984
$e_{600-800}$	0.0017	0.0332	0.0027	0.0434	0.0036	0.0492
$e_{800-900}$	0.0020	0.0362	0.0024	0.0484	0.0052	0.0596
$e_{900-920}$	0.0023	0.0374	0.0033	0.0454	0.0033	0.0459
$e_{920-940}$	0.0032	0.0436	0.0035	0.0458	0.0088	0.0708
$e_{940-960}$	0.0037	0.0464	0.0046	0.0573	0.0044	0.0464
$e_{960-980}$	0.0057	0.0523	0.0062	0.0621	0.0060	0.0594
average	0.0029	0.0403	0.0035	0.0485	0.0062	0.0614

IV. CONCLUSION

In this study, the staked CNN-LSTM prognostic models have been applied, which are combined with two methods, CNN and LSTM. Features are extracted with the original feature sets through convolutional layers and fed to LSTM layers. Finally, an improved recurrent neural network with LSTM layer, which gets adequate consideration of the entire bearing run time and continuous variation, is constructed to predict the bearing performance degradation. The prediction performance is verified using the prognostic data repository of NASA. Furthermore, the comparison with the general CNN and LSTM proves that the stacked convolutional LSTM networks proposed in the paper has an ability of improving prediction performance of the rolling bearing degradation trend.

In future work, it would be necessary to verify, whether the essential features of the vibration waveforms would be extracted correctly through the applied method, and how it changes as training progress.

ACKNOWLEDGMENT

This research was funded by National Natural Science Foundation of China under Gant No. 51875032, 51475098 and 61463010) and Guangxi Natural Science Foundation (2016GXNSFFA380008).

REFERENCE

[1] Biao Wang, Yaguo Lei, Naipeng Li, and J. Lin, "An improved fusion prognostics method for remaining useful life prediction of bearings"[C]//2017 IEEE International Conference on Prognostics and Health Management (ICPHM). IEEE, 2017

[2] Y. Wang, H. Li, J. Yang, D. Yao, C. Li, and J. V. de Oliveira, "Sparse coding based RUL prediction and its application on roller bearing prognostics"[J]. *Journal of Intelligent & Fuzzy Systems* 34.6 (2018): 3719-3733.

[3] Cubillo A, Perinpanayagam S, Esperon-Miguez M. "A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery"[J]. *Advances in Mechanical Engineering*, vol. 8, no. 8, 2016.

[4] Yang W A, Xiao M, Zhou W, et al. "A hybrid prognostic approach for remaining useful life prediction of lithium-ion batteries"[J]. *Shock and Vibration,* vol. 2016, pp. 1-15, 2016.

[5] Tsui K L, Chen N, Zhou Q, et al. "Prognostics and health management: A review on data driven approaches"[J]. *Mathematical Problems in Engineering,* vol. 2015, pp. 1-17, 2015.

[6] Liu J, Wang W, Ma F, et al. "A data-model-fusion prognostic framework for dynamic system state forecasting"[J]. *Engineering Applications of Artificial Intelligence,* vol. 25, no. 4, pp. 814-823, 2012.

[7] Yang J, Nguyen M N, San P P, et al."Deep Convolutional Neural Networks On Multichannel Time Series For Human Activity Recognition," *Twenty-Fourth International Joint Conference on Artificial Intelligence*. 2015.

[8] Yan J, Koc M, Lee J, "A prognostic algorithm for machine performance assessment and its application," *Production Planning & Control* 15.8 (2004): 796-801.

[9] Hinchi A Z, Tkiouat M, "Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network," *Procedia Computer Science* 127 (2018): 123-132.

[10] Gang Tang, Youguang Zhou, Huaqing Wang, and G. Li, "Prediction of Bearing Performance Degradation with Bottleneck Feature based on LSTM Network," *IEEE Instrumentation and Measurement Society,* 14-17 May 2018 2018.

[11] Ordóñez F, Roggen D, "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition," *Sensors* 16.1 (2016): 115

[12] YANN LECUN, I. MEMBER, L ´EON BOTTOU, YOSHUA BENGIO, et al, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE* 86.11 (1998): 2278-2324.

[13] Ming Zeng *et al.*, "Convolutional Neural Networks for human activity recognition using mobile sensors," /6th International Conference on Mobile Computing, Applications and Services. IEEE, 2014: 197-205

[14] Hochreiter S, Schmidhuber J, "Long short-term memory," *Neural computation* 9.8 (1997): 1735-1780.

[15] Kawakami K, *Supervised Sequence Labelling with Recurrent Neural Networks*. *Ph. D. thesis* (2008).

[16] Greff K, Srivastava R K, Koutník J, et al, "LSTM: A Search Space Odyssey," *I IEEE transactions on neural networks and learning systems* 28.10 (2016): 2222-2232.

[17] *The Website of NASA,* the Prognostic Data Repository of bearings is available: https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/

[18] Qiu H, Lee J, Lin J, et al, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of sound and vibration* 289.4-5 (2006): 1066-1090.