Roller bearing degradation assessment based on a deep MLP convolution neural network considering outlier regions

Zhang, Dingcheng; Stewart, Edward; Ye, Jiaqi; Entezami, Mani; Roberts, Clive

DOI:
10.1109/TIM.2019.2929669

License:
Other (please specify with Rights Statement)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Zhang, D, Stewart, E, Ye, J, Entezami, M & Roberts, C 2019, 'Roller bearing degradation assessment based on a deep MLP convolution neural network considering outlier regions', *IEEE Transactions on Instrumentation and Measurement*. https://doi.org/10.1109/TIM.2019.2929669

Roller Bearing Degradation Assessment Based on a Deep MLP Convolution Neural Network Considering Outlier Regions

Dingcheng Zhang, Edward Stewart, Jiaqi Ye, Mani Entezami and Clive Roberts

***Abstract*—Roller bearings are one of the most safety-critical components in many machines. Predicting the vibration-based remaining useful life (RUL) of roller bearings allows operators to make informed maintenance decisions and to guarantee reliability and safety. The health indices (HIs) for degradation assessment are constructed by extracting feature information from the collected data, which significantly influences the prognosis result. Conventional HI construction methods rely heavily on expert knowledge and also have limited capacity for learning health information from the raw data from roller bearings. Furthermore, outlier regions often occur in HIs developed by those methods, and these can easily result in false alarms. To address these problems, a novel HI construction method based on a deep multilayer perceptron (MLP) convolution neural network (DMLPCNN) model, which also considers outlier regions, is proposed in this paper. In the proposed model, a 1-D MLP convolution (Mlpconv) block, consisting of a convolution layer and a micro network, is applied to learn features directly from vibrational data. The learned features are then mapped into an HI using a global average pooling layer and a logistic regression layer. Finally, an outlier region correction method, based on sliding thresholds, is proposed to detect and remove outliers in the HI. The outlier region correction method is able to enhance the interpretability of the constructed HI. The effectiveness of the proposed method is verified using whole-life datasets of 17 bearings. The experimental results demonstrate that the proposed method outperforms conventional methods.**

***Index Terms*—Roller bearings, RUL prediction, degradation assessment, deep MLP convolution neural network, outlier region correction method.**

I. Introduction

Roller bearings are a key component of many machines that potentially operate under heavy loads and changing speed conditions. Faults occurring in roller bearings can result in economic loss or even casualties. The health prognostic of a roller bearing can predict its remaining useful life (RUL) to make an optimal maintenance decision based on the bearing's real operation condition and then avoid the occurrence of tragedies.

This work was supported by the China Scholarship Council. *(Corresponding author: Jiaqi Ye.)*

The authors are with the Department of Electronic, Electrical and Systems Engineering, University of Birmingham, Birmingham B15 2TT, UK. (e-mail: railcm@contact.bham.ac.uk).

Machinery health prognosis has three stages as shown in Fig. 1. Degradation assessment plays a significant role in the health prognosis of roller bearings through constructing a health index (HI). HI construction can identify and quantify a history and ongoing degradation process by extracting operational performance information from the collected data [1, 2]. The quality of the constructed HI directly impacts the accuracy of the health prognostic. HI construction plays a significant role in maintenance and is thus an area of interest for a number of researchers [3].

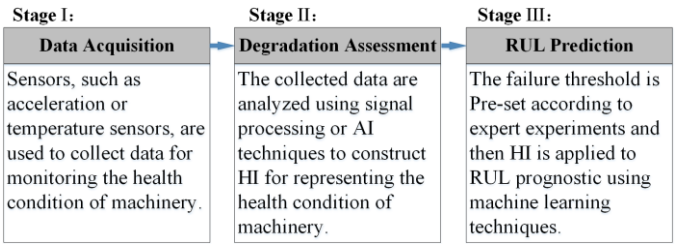


Fig. 1. Three stages of machinery health prognosis.

Recent HI construction methods can be classified into two kinds: physics-based methods and data-driven methods [1]. In physics-based methods, mathematical or physical models are built according to a system's underlying physics, mechanical damage and expert experience. However, the precise nature of the degradation process is difficult to determine due to system and operational complexity, and so it is difficult to build an accurate model. Data-driven methods make use of condition monitoring data to construct HIs without much prior knowledge of the system. Thus, these methods have been widely researched for assessment of bearing degradation. For example, the self-organizing map (SOM) method [4, 5], the principal component analysis method [6, 7] and the Mahalanobis distance [8, 9] were all introduced to fuse multiple features into an HI to be used to assess degradation.

Although the methods listed above demonstrate good performance, they still have some deficiencies. For instance, the performance of the constructed HI often relies heavily on the features selected. Feature selection is a manual procedure, which is time-consuming and requires a domain expert. Methods based on manual feature selection are also difficult to generalize. To overcome these problems, deep learning models have been introduced to learn features from raw data automatically [2, 10–13]. The convolution neural network (CNN) approach, which is one of most used models, has the advantage that it can obtain spatial information from input data. Turker et al. [14] introduced a 1-D CNN to learn features from motor current signals, and

bearing faults were ultimately detected successfully. 1-D CNN has also been used to learn bearing features from vibration signals and then construct an HI in [15]. CNN has demonstrated its suitability for use in condition monitoring in many other target domains [16–19].

The conventional CNN model is constructed by stacking linear convolution blocks, including the convolutional layer and the pooling layer, as shown in Fig. 2 (a). The feature vectors can be extracted using linear convolutional filters and nonlinear activation functions, such as sigmoid, tanh, etc. Thus, the conventional CNN can work well if the latent concepts in the analysis signal are linearly separable. However, the abstract representations which are used to identify the health status of a bearing are highly nonlinear functions. Also, the multiple fully connected layers in the conventional CNN model can easily result in overfitting. Lin et al. [20] proposed a Network in Network (NiN) model which is constructed by stacking Mlpconv (MLP convolution) blocks as shown in Fig. 2 (b). The NiN approach replaces the pooling layer and the fully connected layer in a CNN with a micro network structure and a global average pooling layer, respectively. Inspired by the NiN model, the deep multilayer perceptron (MLP) convolution neural network (DMLPCNN) model is proposed in this paper for the construction of HIs for roller bearings. Significantly, compared to the NiN method, the DMLPCNN method is an "end-to-end" regression model and uses an additional final logistic regression layer. In addition, this model is directly applied to 1-D vibration signals rather than 2-D datasets. The effectiveness of the proposed model is demonstrated with real sensor data.

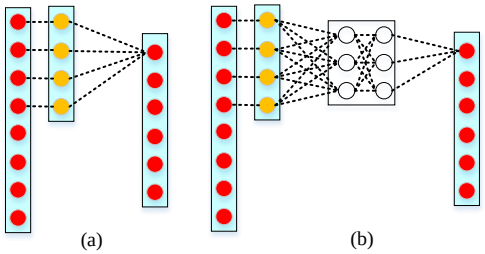


Fig. 2. (a) Linear convolution block structure, (b) Mlpconv block structure.

In addition, another problem for constructed HIs is that some outliers exist. In these HIs, performance for degradation assessment can be significantly affected. To address this problem, the outliers in HIs need to be identified and then removed. Recently, many techniques have been developed to detect outlier regions, including: machine learning-based methods, information theory-based methods and statistical methods [21–23]. The statistical methods are often shown to be both effective and efficient [24]. In particular, the 3σ (3 standard deviation)-based statistical method is widely applied to outlier detection tasks [23, 25]. For example, Guo et al. [15] proposed an outlier region correction method for bearing HIs using the 3σ-based statistical method and demonstrated good results. In that method, the threshold for the HI is fixed and hence outliers with minor amplitudes may be neglected. Additionally, outliers with minor amplitudes or outliers in HIs with nonlinear trends are often not detected when using that approach.

In order to automatically construct HIs, and then to correct outlier regions, this paper proposes an HI construction method based on a DMLPCNN model which also takes into consideration the removal of outliers. In the proposed method, the bearing features are learned using multiple Mlpconv blocks, and the features obtained are then mapped into an HI using both a global average pooling layer and a logistic regression layer. A novel outlier region correction method is then applied to remove outliers in the HIs. The HI based on the proposed method is referred to as an MHI (Mlpconv HI). Run-to-failure datasets from bearings are used to verify the effectiveness of the proposed method. Compared with conventional methods, the proposed method demonstrates advantages when considering evaluation indices associated with the HIs produced. In summary, the main contributions of this paper are:

- The combination of Mlpconv blocks and a global average pooling layer with an additional final logistic regression layer to form an end-to-end regression model (DMLPCNN). The model is then used to construct a bearing HI by learning directly from raw vibration signals.
- Introduction of the use of a stepwise strategy in the application of median-based threshold techniques for a novel outlier removal solution in order to improve the constructed HIs in terms of the evaluation indices used in the work.

This paper is organized as follows: The DMLPCNN model is introduced in Section II. In Section III, a novel outlier region correction method is described. Construction of the evaluation indices for HI comparison is introduced in Section IV. Section V demonstrates the proposed method using the experimental results. Conclusions are then presented in the final section.

II. Deep MLP Convolution Neural Network Model

A novel deep learning model, referred to as the "Deep MLP Convolution Neural Network" (DMLPCNN) model", is proposed in this paper to obtain the HI of roller bearings. The proposed model includes two 1-D Mlpconv blocks, one global average pooling layer and one logistic regression layer, as shown in Fig. 3.

A. 1-D Mlpconv Block

The Mlpconv block includes a convolution layer and a micro network forming an MLP. The micro network is a nonlinear structure that replaces the linear pooling layer in a CNN. Thus, the Mlpconv block can be used to obtain a more abstract representation of the input data than a standard CNN. The main difference between 2-D and 1-D Mlpconv blocks is that 2-D matrices are replaced by 1-D arrays for both kernels and feature maps. The 1-D Mlpconv block can be expressed as [20]:

$$\begin{aligned} f_{i,k_1}^{1} &= \max({\omega_{k_1}^{1}}^{T} x_i + b_{k_1}, 0) \\ &\vdots \\ f_{i,k_n}^{n} &= \max({\omega_{k_n}^{n}}^{T} f_i^{n-1} + b_{k_n}, 0) \end{aligned} \tag{1}$$

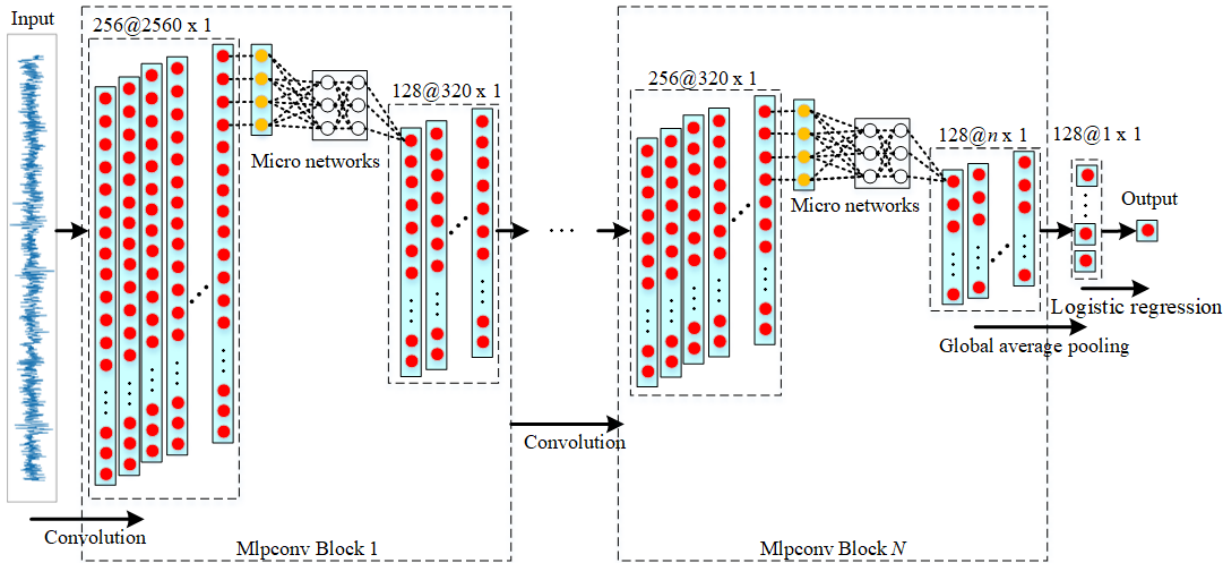


Fig. 3. Deep MLP convolution neural network model.

where i is the index in the feature map, x_i is the input array section at location i, k is the channel number of the feature map, N is the number of layers within the micro networks (selection of N is analyzed in Section V), and $\omega_{k_j}^j$ and b_{k_j} are the weight and the bias at the j-th layer in a micro network ($j = 1, \cdots, n$). In the paper, the activation function in the Mlpconv block is the tanh unit. Equation (1) shows that a parametric pooling process, consisting of a micro network and the pooling layer, is added following a normal convolution layer. The parametric pooling structure is able to obtain complex and learnable interactions of information.

B. *Global Average Pooling Layer*

In a conventional CNN, abstract feature maps are obtained through multiple combinations of the convolution layer and the pooling layer. The obtained feature maps are then fed into a series of fully connected layers followed by a logistic regression layer for classification or prediction. However, use of the fully connected layers can result in overfitting and thus reduce the generalization capabilities of the deep learning model [20]. Furthermore, the number of parameters in the fully connected layers is normally too high, leading to a time-consuming optimization procedure. Hence, a global average pooling layer is added following the Mlpconv blocks in the proposed model. The global average layer returns the average value of the last Mlpconv block's result for each channel as follows [20]:

$$p_k = \text{Ave}\{f_{i,k_2}^2\} \tag{2}$$

where p_k is the output of the k-th channel feature map, Ave{■} represents the average operation, and f_{i,k_2}^2 is the result of the second Mlpconv block.

C. *Logistic Regression Layer*

After using the global average layer, the output for each channel, p_k ($k = 1,2, \cdots, n$), can be obtained as demonstrated in (2). The combination of results from the global average pooling layers from all channels are given as P in (3). To ensure the value of the constructed HI is in the range 0 to 1, the last layer of the proposed model is a logistic regression layer, as described in (4).

$$P = [p_1, p_2, \cdots, p_n] \tag{3}$$

$$\bar{y} = 1 - \frac{1}{1 + e^{-(WP+b)}} \tag{4}$$

where W and b are the weights and bias, respectively, and $\bar{y}$ is the result of the DMLPCNN model, also known as the constructed HI. To optimize the parameters in the proposed model, an objective function, J, is constructed as in [15]:

$$\text{J} = \sum_{j=1}^{N} \|y_j - \bar{y}_j\|_1^2 \tag{5}$$

where N is the number of training samples, and y_j and $\bar{y}_j$ are the actual label and the expected label, respectively, for the j-th training sample. Optimal values for the parameters in the proposed model (to minimize error in label allocation) can be obtained by minimizing (5).

III. Outlier Region Correction Method

Although an HI can be constructed for a target bearing using the proposed model, as each new estimate of HI is based on an incremental variation from the previous estimate, any outliers in the HI would have a cumulative effect that may potentially result in a false alarm being generated. In order to remove the HI outliers, a novel outlier region correction method is proposed in this paper. Unlike conventional HI outlier removal methods, such as the 3σ-based method described in [15], a stepwise strategy is first applied to divide the analyzed HI into

multiple sections. A robust threshold based on the median is then calculated (as per [26]) for each section in order to detect outlier regions. Hence, the outliers can be removed.

The difference between *HI* points, *dHI*, is:

$$dHI^k = \frac{HI^{k+1} - HI^k}{\Delta t} \tag{6}$$

where HI^k and HI^{k+1} are the *k*-th and (*k*+1)-th points in the HI, respectively. Δt is the time interval between HI^k and HI^{k+1}. A sliding window with fixed length, *L*, is used to divide *dHI* into multiple sections. The overlap rate, α, is set to avoid missing outliers. The start point, ps_i, and the end point, pe_i, of the *i*-th section of *dHI* are shown as:

$$\begin{cases} ps_i = (1-\alpha)iL \\ pe_i = ps_i + L \end{cases} \tag{7}$$

A common threshold is constructed as $\mu \pm 3\sigma$. This form is widely used to detect outliers, where μ and σ are the mean and standard deviation, respectively. The value of the mean is more sensitive to outliers than that of the median. To construct a robust threshold, (8) and (9) are used, as introduced in [26].

$$MAD_i = \text{med}(|dHI_i - \text{med}(dHI_i)|) \tag{8}$$

$$Threshold_i = \begin{cases} med_i + \beta \times MAD_i \\ med_i - \beta \times MAD_i \end{cases} \tag{9}$$

where med(■) is the median operation, dHI_i is the difference of *HI* in the *i*-th section, and med_i and MAD_i are the median and median absolute deviation of dHI_i, respectively. β is an HI performance parameter which is identified by analysis and is described in more detail in Section V.

The points of *dHI* whose values are greater than the upper threshold are referred to as positive outliers; points of *dHI* whose values are less than the lower threshold are referred to as negative outliers. If a localized region of *dHI* points includes *l* consecutive positive outliers, and *l* consecutive negative outliers, the region is considered to be an outlier region and is removed from the HI in its entirety. Based on work described in [15], *l* is selected as 5.

To evaluate the performance of the proposed method, a simulation experiment has been conducted with the outputs shown in Fig. 4. Fig. 4 (a) and (c) show linear and nonlinear HIs with three outlier regions (O_1, O_2, O_3). Fig. 4 (b) and (d) show the incremental differences in the two HIs. Fig. 4 (b) indicates that O_2 cannot be detected by the 3σ-based method because of the interference of the other outliers. However, the proposed method identifies the smaller outlier effectively. Fig. 4 (d) shows that the 3σ-based method fails to identify the outlier regions due to the nonlinearity of the HI, but that the proposed method successfully identifies them.

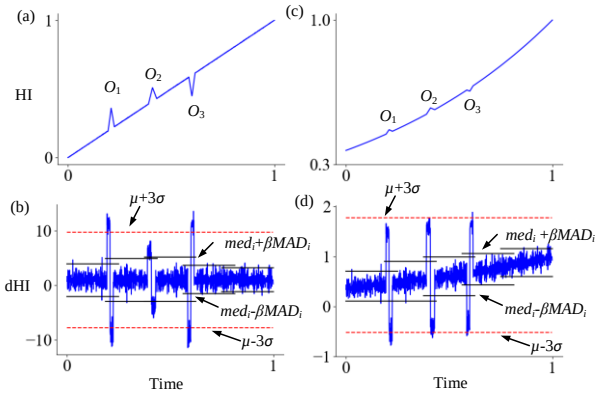


Fig. 4. (a) Linear HI with outliers, (b) difference of HI in (a), (c) nonlinear HI with outliers, (d) difference of HI in (c).

After detection of the outlier regions, (10) is introduced to remove the outliers [15].

$$MHI_{to} = HI_{ts} + \frac{HI_{te} - HI_{ts}}{te - ts}(to - ts) \tag{10}$$

where *to* is the time in the outlier region, and *ts* and *te* are the start time and end time, respectively, of the outlier region. The details of the outlier region correction method are demonstrated in Algorithm 1, as shown below.

Algorithm 1 outlier region correction.

Input: health index (*HI*), length of HI (*K*), length of slide windows (*L*), overlapping rate for sub-sections (α), threshold parameter (β).
Initialization: numbers of negative and positive outliers in one region n_n and n_p, start and end points ps_1 and pe_1.
Output: MHI
Procedures:
1. Compute the difference of HI, *dHI*, according to (6)
2. **While** $pe_i \leq K$-1 do
3. Compute the threshold in *i*-th section through (8) and (9)
4. **for** *j*: = ps_i to pe_i do
5. **if** dHI^j < the lower threshold, then $n_{n,i} = n_{n,i} + 1$
6. **else if** dHI^j > the upper threshold, then $n_{p,i} = n_{p,i} + 1$
7. **end if**
8. **end for**
9. **if** $n_{n,i} \geq l$ and $n_{p,i} \geq l$, then correct the outlier region through (10)
10. **end if**
11. $i = i + 1$
12. Compute the start and the end points of the *i*-th section, ps_i and pe_i, through (7)
13. **end while**

IV. Evaluation Indices for HI Construction

To assess the quality of the constructed HIs, three evaluation indices are introduced in [1]: monotonicity, robustness and trendability. The monotonicity index, demonstrated in (11), can be used to evaluate the monotonicity of the HI, i.e. its nature in respect of a continuously increasing or decreasing value. This kind of trend would be expected in an HI as the degradation process is irreversible. The robustness index, shown in (12), is designed to assess the stability of the HI. This is essentially a quantification of noise, stochasticity of the degradation process,

and other random fluctuations in the HI. Finally, given that roller bearings usually degrade over time and with use, trendability, as described as (13), is used to evaluate the correlation between the degradation trend of an HI and the time of operation.

$$Mon(HI) = \frac{|\text{No. of } dHI > 0 - \text{No. of } dHI < 0|}{K-1} \quad (11)$$

$$Rob(HI) = \frac{1}{K}\sum_{k=1}^{K} \exp\left(-\left|\frac{HI_k - HI_k^T}{HI_k}\right|\right) \quad (12)$$

$$Tre(HI,t) = \frac{|\sum_{k=1}^{K}(HI_k - \overline{HI})(t_k - \bar{t})|}{\sqrt{\sum_{k=1}^{K}(HI_k - \overline{HI})^2 \sum_{k=1}^{K}(t_k - \bar{t})^2}} \quad (13)$$

where K is the length of the HI, No. of $dHI > 0$ and No. of $dHI < 0$ are the number of positive and negative differences, respectively, HI_k is the value of the HI at time t_k, and HI_k^T is the mean trend value of the HI at t_k which is obtained by using an average smoothing method, $\overline{HI} = (\sum_{k=1}^{K} HI_k)/K$, $\bar{t} = (\sum_{k=1}^{K} t_k)/K$.

The three evaluation indices above assess different properties of the HI. In order to comprehensively evaluate an HI, a composite index (CI) is constructed as:

$$CI = \frac{Mon + Rob + Tre}{3} \quad (14)$$

V. Experiment and Discussion

A. Data Description

Vibrational datasets of roller bearings made openly available in [27] were used to verify the effectiveness of the proposed method. The 17 vibrational datasets were collected from an experimental platform known as PRONOSTIA, as shown in Fig. 5. The platform which is composed of a rotating part, a degradation generation part and a measurement part allows bearing degradations to be conducted in only a few hours. The parameters of the test bearings are shown in Table I. In the data collection procedure, samples were recorded every 10 s until failure occurred. Datasets were collected from different bearings operating under three different speed and load conditions, as shown in Table II. Information relating to the operating conditions is not made available to the processing in order that the robustness of the method can be demonstrated. Sixteen sets of data corresponding to test bearings were randomly selected and used to construct the training dataset; the remaining one was used as the testing dataset. It is worth noting that samples are normalized before being input to the proposed model in order to improve the training efficiency and accuracy. Thus, the values of mean and standard deviation for the normalized input samples are 0 and 1, respectively. For ease of manipulation, the run-to-failure dataset is labeled ranging from 0 to 1. The first sample (normal operation) is therefore labeled as 0 and the last sample (failed status) is labeled as 1. Intermediate labels are generated by dividing the time of operation by the whole-life time of the bearing. This linear approximation is a valid starting point for HI generation using data-driven approaches [28, 29] and is required as the precise bearing degradation curve is not known.

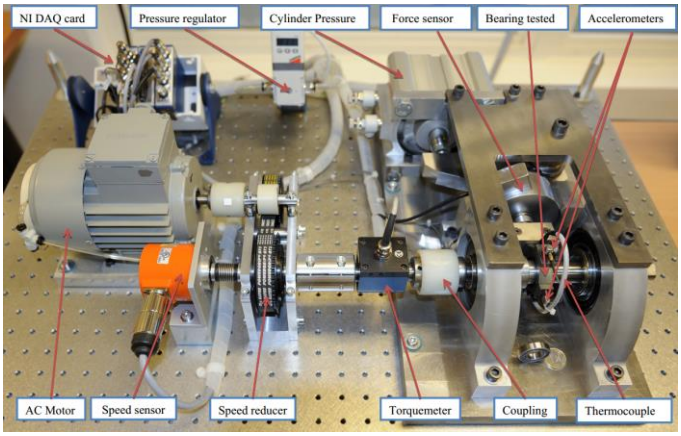


Fig. 5. Experimentation platform for the test.

TABLE I
PARAMETERS OF TEST BEARINGS

Outer race diameter	Inner race diameter	Roller diameter	Roller number
32 mm	22.1 mm	3.5 mm	13

TABLE II
OPERATING CONDITIONS FOR BEARINGS

Bearing Number	Speed	Load
1–7	1800 rpm	4000 N
8–14	1650 rpm	4200 N
15–17	1500 rpm	5000 N

B. HI Construction

The number of Mlpconv blocks, *N*, determines the structure of the proposed DMLPCNN model. To identify the appropriate value for *N*, a comparison experiment using different *N* values has been conducted. The results are summarized in Fig. 6. The figure shows that above 30 epochs, the loss function for the testing dataset is minimized with 2 blocks. As the model does not converge until ~60 epochs and is run to 100 epochs, this becomes the significant portion of the curve. Hence, the number of Mlpconv blocks used in this work is 2.

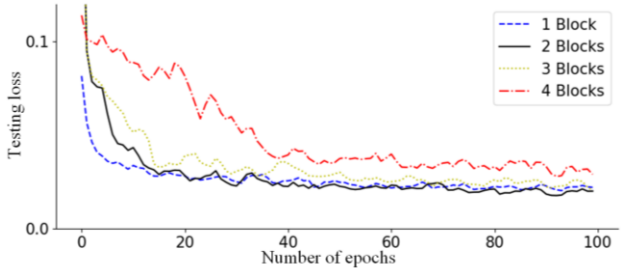


Fig. 6. Loss functions for the testing dataset for different numbers of Mlpconv blocks.

The kernel length and the width of the micro network are critical hyper-parameters in the proposed model, and directly influence the HI construction. To select the optimal value of kernel length, a comparison experiment is conducted using the proposed method with different kernel lengths. Fig. 7 is a boxplot showing the results of this comparison experiment. The figure shows that the median values of CI at kernel lengths of 15 and 20 are similar but the CIs are more centralized at 20. Hence, the best performance is obtained when a kernel length of 20 is selected. A similar comparison experiment is used to select the width of the micro network with the result shown in Fig. 8. The figure shows that median values at widths of 16 and 32 are similar; however, the distribution of samples is more concentrated at 32. Furthermore, the CI values at other widths have outliers. Also, the time costs for different micro network widths in each epoch, *te*, are shown in Table III. The time cost for the calculation required to update the model increases as the width of the micro network is increased, but the difference in time cost between widths 16 and 32 is proportionally small. Thus, the width of the micro network is selected as 32 in the proposed model.

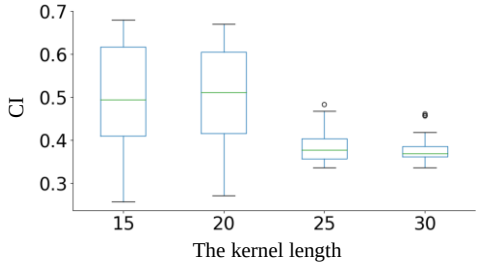


Fig. 7. Composite index results for different kernel lengths.

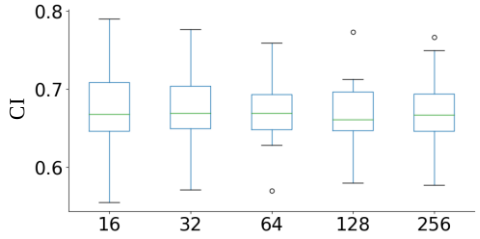


Fig. 8. Composite index results for different widths of micro network.

TABLE III
TIME COST FOR DIFFERENT MICRO NETWORK WIDTHS

Width	16	32	64	128	256
te	1.08 s	1.09 s	1.17 s	1.35 s	1.71 s

With the critical parameters selected, the proposed model can be used to construct HIs. Fig. 9 shows example HIs for the 5th and 6th bearings, respectively. The figure shows that the amplitude of the HIs generally increases with the time of operation until it reaches the failure point at a value of 1. However, some noise exists in the His, as shown in Fig. 9 (a) and (c). A moving average algorithm is applied to smooth the His, with results shown in Fig. 9 (b) and (d). This smoothed version is used in further processing.

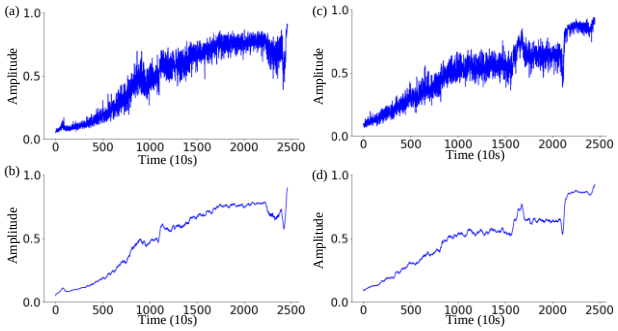


Fig. 9. (a) and (b) HIs for the 5th bearing before and after smoothing, (c) and (d) HIs for the 6th bearing before and after smoothing.

C. Outlier Region Correction

HIs are generally monotonous. Thus, the outlier region correction method proposed in Section III can be applied. The result of the proposed method can, however, be influenced by several parameters, including the length of slide windows, L, the crossover rate, α, and the threshold parameter, β, in (9). Comparison experiments are also used to select appropriate values for these parameters.

The effects of different slide window sizes and crossover rates are demonstrated using the boxplots shown in Fig. 10. The crossover rates in Fig. 10 (a), (b) and (c) are selected as 0.50, 0.55 and 0.60, respectively. In the figure, the maximum medians of CI are at a length of 150, 150 and 200 when α is 0.5, 0.55 and 0.6, respectively. However, considering minimizing outliers of the CI, the slide window length and the crossover rate are selected as 150 and 0.55, respectively. An additional comparison experiment is conducted to select the optimal threshold parameter, β. The result, displayed in Fig. 11, shows that the best performance is obtained when β is set to 0.3.

Using the selected parameters, MLP-based health indices (MHIs) for the example bearings can be obtained using the proposed method. In Fig. 12, the red dashed line and the blue solid line represent HIs and MHIs from bearing 1 to bearing 9, respectively. The figure shows that some outliers in the HIs are removed by using the proposed method and thus MHIs have better interpretability. Fig. 13 shows the composite indices of HIs and MHIs for all of the test bearings. The figure demonstrates that MHIs have better performance than HIs in terms of the overall CI, which verifies the effectiveness of the proposed method. It should be noted that the CI is constructed from three elements, monotonicity, robustness and trendability, and as such is highly dependent on the shape of the HI curve. All bearings tested are run to failure, but the shapes of the HI curves vary, as shown in Fig. 12.

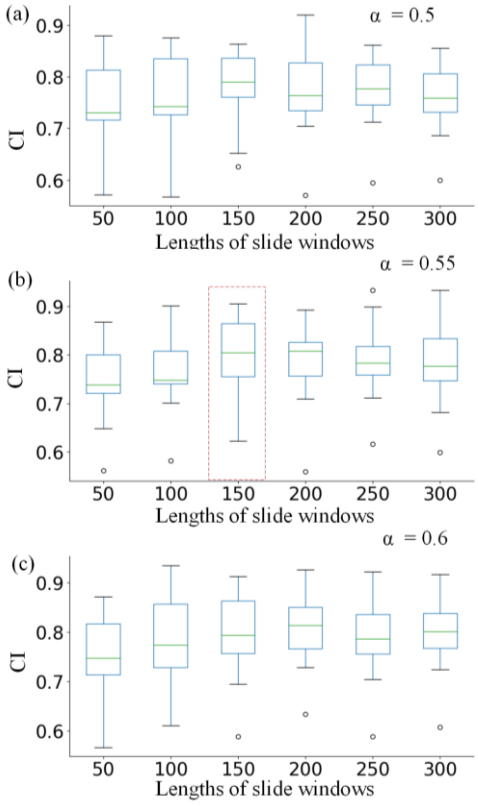


Fig. 10. Composite index results for different slide window lengths and crossover rates. (a) Crossover rate: 0.50, (b) crossover rate: 0.55, and (c) crossover rate: 0.60.

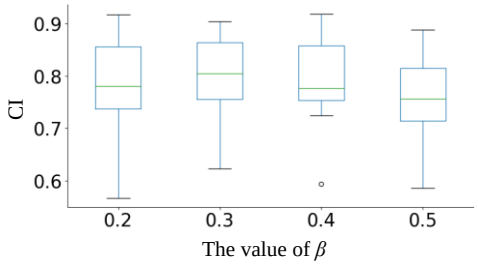


Fig. 11. Composite index results for the proposed outlier region correction method with different values of the threshold parameter, β.

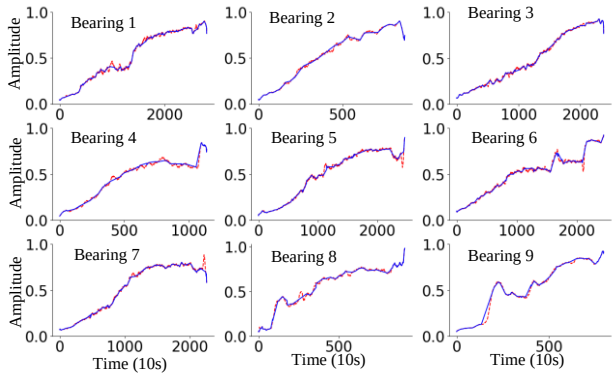


Fig. 12. HIs (red/dashed) and MHIs (blue/solid) from bearing 1 to bearing 9.

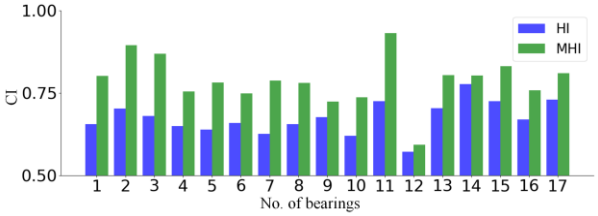


Fig. 13. Composite indices of HIs and MHIs for all test bearings.

D. *Comparison*

Overfitting is a common problem in machine learning. When overfitting occurs, the trained model will align well with the training dataset, but poorly with the testing dataset. Fig. 14 (a) and (b) demonstrate results in terms of error between the true and predicted labels (J, in (5)) obtained using both the conventional CNN [15] and the DMLPCNN models. This error is referred to as "loss". In the figure, the dashed line and the solid line display the changing trends of the training loss and the testing loss, respectively. Fig. 14 (a) shows that the gap between the training loss and the testing loss occurs at approximately the 15th epoch and then becomes greater as the epoch number (number of training events) increases. It also indicates that overfitting occurs when using the conventional CNN model, but not when using the proposed DMLPCNN model.

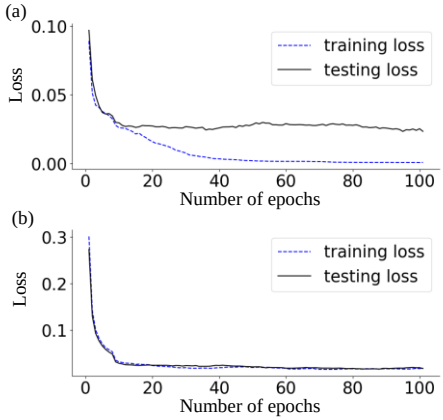


Fig. 14. Loss function in relation to epoch: (a) conventional CNN model, (b) DMLPCNN model.

To show the advantage of the proposed model, six other HI construction methods were conducted for comparison. In the first method, a deep learning model, a stacked autoencoder (SAE) with three hidden layers, is used to construct HIs, in which the input is the raw vibrational datasets. The second uses the SOM method and applies it to specifically designed features in order to obtain HIs [5]. In the third one, a fully connected (FC) neural network based on handcrafted features is utilized to construct HIs. The fourth method uses the conventional CNN model to obtain HIs [15]. Similar to the method in [15], two advanced versions of CNN, i.e. fully convolutional network [30] and residual network (ResNet) [31], are also applied to construct HIs for the final two comparison cases. The results of the comparisons in terms of indices are shown in Table IV, which demonstrates the obvious advantage of the DMLPCNN method.

TABLE IV
COMPARISON RESULTS FOR DIFFERENT HI CONSTRUCTION METHODS

Method	Mon	Rob	Tre	CI
SAE	0.02	0.10	0.12	0.08
SOM	0.15	0.35	0.80	0.43
FC	0.23	0.41	0.79	0.48
CNN	0.40	0.61	0.89	0.63
FCN	0.31	0.53	0.80	0.55
ResNet	0.48	0.72	0.92	0.71
DMLPCNN	**0.65**	**0.89**	**0.95**	**0.83**

To further demonstrate the superiority of the proposed outlier removal method, further comparisons with the 3σ-based outlier region correction method [15] were conducted. In the comparison, the HI is constructed by using the DMLPCNN model first, and then the proposed and 3σ-based methods are used to remove outliers. The results of the comparisons in terms of CI are shown in Fig. 15, which indicates that the proposed MHI method is superior in terms of overall HI quality.

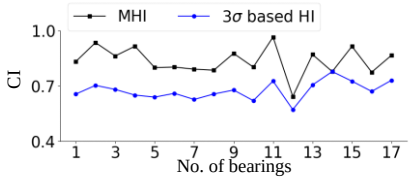


Fig. 15. Composite indices of MHIs and 3σ-based HIs.

VI. CONCLUSION

In this paper, the DMLPCNN model is proposed and used to construct HIs for roller bearings. A novel outlier region correction method is then proposed and applied in order to improve the quality of the constructed HIs. The effectiveness of the proposed combined method has been verified using comparative studies with publicly accessible run-to-failure datasets for example bearings. Compared with conventional methods, the proposed method has been shown to generate HIs with greater overall quality, as demonstrated through a CI made up of three key performance metrics. The following specific conclusions are drawn following consideration of the comparison experiments:

(1) The 1-D Mlpconv block in the DMLPCNN model is able to obtain abstract features from bearing data with little prior information. The HIs constructed using the proposed model have good overall quality as indicated using a number of evaluation indices. In addition, the application of a global average pooling layer effectively avoids the problem of overfitting.

(2) In the novel outlier region correction method, the use of sliding thresholds based on the median and median absolute deviation are more robust than standard thresholds based on mean and standard deviation. The proposed method has been shown to cope with minor outliers and also outliers in HIs with nonlinear behavior.

(3) The overall effectiveness of the proposed method has been demonstrated using publicly accessible datasets for roller bearings collected using a test rig. The authors would next like to consider application of the technique to data collected from a real operating environment.

REFERENCES

[1] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018.

[2] M. Ma, C. Sun, and X. Chen, "Discriminative deep belief networks with ant colony optimization for health status assessment of machine," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, pp. 3115–3125, 2017.

[3] L. Cui, X. Wang, Y. Xu, H. Jiang, and J. Zhou, "A novel switching unscented Kalman filter method for remaining useful life prediction of rolling bearing," *Measurement*, vol. 135, pp. 678–684, 2019.

[4] H. Qiu, J. Lee, J. Lin, and G. Yu, "Robust performance degradation assessment methods for enhanced rolling element bearing prognostics," *Advanced Engineering Informatics*, vol. 17, pp. 127–140, 2003.

[5] Y. Zhang, B. Tang, Y. Han, and L. Deng, "Bearing performance degradation assessment based on time-frequency code features and SOM network," *Measurement Science and Technology*, vol. 28, p. 045601, 2017.

[6] A. Widodo and B.-S. Yang, "Application of relevance vector machine and survival probability to machine degradation assessment," *Expert Systems with Applications*, vol. 38, pp. 2592–2599, 2011.

[7] S. Dong and T. Luo, "Bearing degradation process prediction based on the PCA and optimized LS-SVM model," *Measurement*, vol. 46, pp. 3143–3152, 2013.

[8] X. Jin, Y. Sun, Z. Que, Y. Wang, and T. W. Chow, "Anomaly detection and fault prognosis for bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, pp. 2046–2054, 2016.

[9] J. Yu, "Health condition monitoring of machines based on hidden Markov model and contribution analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, pp. 2200–2211, 2012.

[10] H. Shao, H. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mechanical Systems and Signal Processing*, vol. 102, pp. 278–297, 2018.

[11] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, pp. 185–195, 2018.

[12] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, pp. 1693–1702, 2017.

[13] W. Lu, Y. Li, Y. Cheng, D. Meng, B. Liang, and P. Zhou, "Early fault detection approach with deep architectures," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, pp. 1679–1689, 2018.

[14] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Transactions on Industrial Electronics*, vol. 63, pp. 7067–7075, 2016.

[15] L. Guo, Y. Lei, N. Li, T. Yan, and N. Li, "Machinery health indicator construction based on convolutional neural networks considering trend burr," *Neurocomputing*, vol. 292, pp. 142–150, 2018.

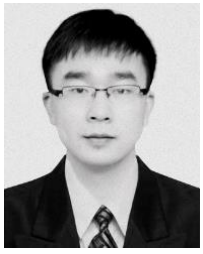
[16] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, and D. J. Inman, "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks," *Journal of Sound and Vibration*, vol. 388, pp. 154–170, 2017.

[17] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, *et al.*, "Convolutional neural network based fault detection for rotating machinery," *Journal of Sound and Vibration*, vol. 377, pp. 331–345, 2016.

[18] J. Feng, F. Li, S. Lu, J. Liu, and D. Ma, "Injurious or noninjurious defect identification from MFL images in pipeline inspection using convolutional neural network," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, pp. 1883–1892, 2017.

[19] H. Wang, S. Li, L. Song, and L. Cui, "A novel convolutional neural network based fault recognition method via image fusion of multi-

vibration-signals," *Computers in Industry*, vol. 105, pp. 182–190, 2019.
[20] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint*, arXiv:1312.4400, 2013.
[21] M. Rahmani and G. K. Atia, "Randomized robust subspace recovery and outlier detection for high dimensional data matrices," *IEEE Transactions on Signal Processing*, vol. 65, pp. 1580–1594, 2017.
[22] S. Çınar and N. Acır, "A novel system for automatic removal of ocular artefacts in EEG by using outlier detection methods and independent component analysis," *Expert Systems with Applications*, vol. 68, pp. 36–44, 2017.
[23] H. Tan, J. A. Maldjian, J. M. Pollock, J. H. Burdette, L. Y. Yang, A. R. Deibler, *et al.*, "A fast, effective filtering method for improving clinical pulsed arterial spin labeling MRI," *Journal of Magnetic Resonance Imaging*, vol. 29, pp. 1134–1139, 2009.
[24] V. Chandola, A. Banerjee, and V. Kumar, "Outlier detection: A survey," *ACM Computing Surveys*, vol. 41, 2007.
[25] N. Li, Y. Lei, J. Lin, and S. X. Ding, "An improved exponential model for predicting remaining useful life of rolling element bearings," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 7762–7773, 2015.
[26] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *Journal of the American Statistical Association*, vol. 88, pp. 1273–1283, 1993.
[27] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, *et al.*, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in *IEEE International Conference on Prognostics and Health Management, PHM'12*, 2012, pp. 1–8.
[28] J. Deutsch and D. He, "Using deep learning-based approach to predict remaining useful life of rotating components," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, pp. 11–20, 2018.
[29] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliability Engineering & System Safety*, vol. 172, pp. 1–11, 2018.
[30] R. Yao, G. Lin, Q. Shi, and D. C. Ranasinghe, "Efficient dense labelling of human activity sequences from wearables using fully convolutional networks," *Pattern Recognition*, vol. 78, pp. 252–266, 2018.
[31] W. Xie, Y. Li, and X. Jia, "Deep convolutional networks with residual learning for accurate spectral-spatial denoising," *Neurocomputing*, vol. 312, pp. 372–381, 2018.

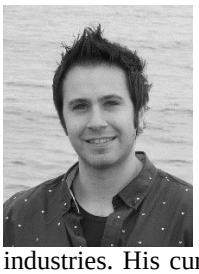


Dingcheng Zhang received the Master's degree in mechanical engineering in 2016 from Hunan University, Hunan, China.

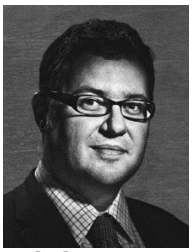
He is currently pursuing the PhD degree in the Department of Electronic, Electrical and Systems Engineering at the University of Birmingham, Birmingham, UK. His research interests include signal processing and machine learning, fault diagnosis and prognosis, intelligent condition monitoring and maintenance.

Edward Stewart (MEng, PhD) is a specialist in railway condition monitoring. Dr Stewart leads the Sensing and Autonomous Systems research theme in the UK Railway Research and Innovation Network, as well as the condition monitoring activities within the Birmingham Centre for Railway Research and Education. His research interests lie in the instrumentation and algorithms for condition monitoring of railway vehicles and fixed railway assets.

Jiaqi Ye (BEng, MRes) has a background in Electronics and Information Engineering, having studied at both Huazhong University of Science and Technology (China) and the University of Birmingham, Birmingham (UK). Jiaqi is now a PhD student at the Birmingham Centre for Railway Research and Education at the University of Birmingham. His research interests include railway components condition monitoring, multi-sensor integration-based railway inspection and signal processing.



Mani Entezami (MSc, PhD) is a specialist in railway condition monitoring. Dr Entezami is a research fellow in the Birmingham Centre for Railway Research and Education and has developed novel techniques and condition monitoring systems that have successfully been deployed in the rail and renewable energy industries. His current research portfolio includes projects on acoustic and vibration signals, low-power and energy-harvesting wireless sensor networks and high-speed embedded systems for data logging and signal processing.



Clive Roberts (M'14) (BEng, PhD) is Professor of Railway Systems at the University of Birmingham and the director of the Birmingham Centre for Railway Research and Education. Prof Roberts leads the UK Rail Research and Innovation Network and the Centre for Excellence in Digital Systems. His research interests include systems engineering, system modeling and simulation, traffic management, fault detection and diagnosis, and data collection and decision support applied to railway traction, signaling, mechanical interactions and capacity.