Deep-Convolution-Based LSTM Network for Remaining Useful Life Prediction

Meng Ma and Zhu Mao

***Abstract*—Accurate prediction of remaining useful life (RUL) has been a critical and challenging problem in the field of prognostics and health management (PHM), which aims to make decisions on which component needs to be replaced when. In this article, a novel deep neural network named convolution-based long short-term memory (CLSTM) network is proposed to predict the RUL of rotating machineries mining the *in situ* vibration data. Different from previous research that simply connects a convolutional neural network (CNN) to a long short-term memory (LSTM) network serially, the proposed network conducts convolutional operation on both the input-to-state and state-to-state transitions of the LSTM, which contains both time–frequency and temporal information of signals, not only preserving the advantages of LSTM, but also incorporating time–frequency features. The convolutional structure in the LSTM has the ability to capture long-term dependencies and extract features from the time–frequency domain at the same time. By stacking the multiple CLSTM layer-by-layer and forming an encoding-forecasting architecture, the deep learning model is established for RUL prediction in this article. Run-to-failure tests on bearings are conducted, and vibration responses are collected. Using the proposed algorithm, RUL is predicted, and as a comparison, the performance from other methods, including deep CNNs and deep LSTM, is evaluated using the same dataset. The comparative study indicates that the proposed CLSTM network outperforms the current deep learning algorithms in URL prediction and system prognosis with respect to better accuracy and computation efficiency.**

***Index Terms*—Convolutional operation, deep learning, long short-term memory (LSTM) network, remaining useful life (RUL) prediction.**

I. Introduction

PROGNOSTICS and health management (PHM) is an emerging technique in mechanical engineering that is gaining interests from both academia and industry [1]. Since unexpected breakdowns will cause lost production, failed shipping schedule, and poor customer satisfaction [2], it is necessary to adopt the PHM technique, which provides users with an integrated view of the health state of a machine or system [3] and is expected to reduce unnecessary maintenance and improve their reliability and safety as well [4]. Accurate prediction of remaining useful life (RUL) after a fault has been detected as a key part of PHM implementation [5]. Once the fault occurs, the RUL estimation needs to be implemented in order for timely maintenance actions to be carried out to avoid catastrophic failures.

In the real world, mechanical systems usually undergo gradual degradation rather than failing abruptly. Therefore, an ideal health indicator is usually constructed first to characterize the trend of the fault progression in a system's whole lifetime [6], and the RUL prediction is carried out based on the established health indicator through different algorithms [7]. In general, RUL prediction methods can be categorized into physics-based, data-driven, and hybrid approaches. Physics-based methods aim to predict the RUL through building a mathematical model, which describes the system's degradation process; therefore, the knowledge of the system's failure mechanism is needed in advance. Daiele and Goebel [8] developed a model-based prognostics algorithm using particle filters and investigated the benefits of the model-based approached with diminishment of sensor sets. In [9], a tool for online prognostic based on a particle filter and optimized tuning kernel smoothing was presented. El Mejdoubi *et al.* [10] proposed a particle-filter-based RUL estimation model considering the aging conditions such as temperature and voltage in the developed degradation law. However, it is undoubtedly difficult to understand and model all failure modes and degraded system dynamics for any complex system under a wide range of working conditions, which constrain applicability and accuracy of the RUL prediction through physics-based modeling.

Hybrid methods, which combine the aforementioned physical models with data-driven models to improve the prediction performance, appear to be a promising solution to tackle the challenging RUL prediction problem. There are a number of implementations using hybrid methods. Sun *et al.* [11] fused the physics-based model of the Wiener process and the data-driven model of empirical mode decomposition and neural network to predict the RUL of cutting tools, and a great potential in such applications is foreseeable in the experimental results. However, designing a fusion mechanism to successfully hybridize the two approaches is not a straightforward task. In addition, most of these methods are unable to be updated with online measured data, and these difficulties limit the potential of the current hybrid

Manuscript received November 6, 2019; revised February 27, 2020; accepted April 12, 2020. Date of publication May 7, 2020; date of current version November 20, 2020. This material is based upon work supported by the Air Force Office of Scientific Research under Award FA95501810491. Any opinions, finding, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Air Force. Paper no. TII-19-4892. *(Corresponding author: Zhu Mao.)*

The authors are with the Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, MA 01854 USA (e-mail: meng_ma@uml.edu; zhu_mao@uml.edu).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2020.2991796

approaches [12]. Therefore, with significant development of sensors, sensor networks, and computing systems, data-driven methods have drawn more and more attention in order to improve the prediction performance.

Recurrent neural networks (RNNs) have shown their advantages in modeling data with a sequential structure and have been used for time-series prediction due to their ability in capturing long-term dependencies of the time history [13]. However, when long short-term memory (LSTM) networks are used for bearing's RUL prediction, statistical features extracted from vibration signals do not well represent the nonstationary information [14], as such transient information is amortized through processing the whole series. In practical bearing's degradation, nonstationary characteristics are crucial to the prognostics. LSTMs oftentimes fail to provide valuable degradation information using extracted statistical features with only time-domain or frequency-domain analysis. To address this problem, a novel convolution-based long short-term memory (CLSTM) network is proposed in this article. The convolutional operation is conducted in the state transitions of LSTM, which is different from the traditional combination of convolution neural network (CNN) and LSTM in [15]. The CLSTM deals with time–frequency features through the embedded convolutional operation. By stacking the CLSTM onto a deep architecture, high-level features may be extracted for RUL prediction through a regression layer. To validate the proposed prognostic algorithm, bearings' run-to-failure test is carried out, and ambient vibration response is collected and used for deep CLSTM training and verification. The results demonstrate a great effectiveness of the proposed method and superiority by comparing with other state-of-the-art methods. The main contributions of this article are summarized as follows.

1) A CLSTM learning algorithm is proposed to extract the temporal time–frequency features, and the transitions of inputs-to-states and states-to-states are modeled through convolutions instead of traditional multiplication. A series of long-term time–frequency features are considered sequentially when predicting RUL.
2) A deep architecture is developed based on the proposed CLSTM to learn high-level features in the time–frequency domain for RUL prediction. The spatial features and sequential features are learned through convolutional operation and temporal modeling, which are more efficient at feature representation through a hierarchical architecture.
3) With bearing run-to-failure test data, the RUL prediction through the constructed deep CLSTM outperforms the benchmark models.

The rest of this article is organized as follows. Section II discusses the related work. Section II presents the proposed model, CLSTM, and the corresponding deep architecture. Section IV introduces the RUL prediction framework and bearings' run-to-failure experiments. Section V presents validation results and discussion compared with other methods. Finally, Section VI concludes this article.

II. Literature Review

Data-driven methods focus on modeling historically collected data to identify current health state and predict RUL. Instead of struggling with the vague knowledge that might be available in characterizing the fatigue and degradation, data-driven approaches employ artificial intelligence to learn the failure path and extract features through processing big volume of data. Chen *et al.* [16] proposed a novel prognostics model based on relative features and multivariable support vector machine for RUL of bearings with small samples. Elforjani and Shanbr [17] adopted three supervised machine learning techniques to correlate acoustic emission features for RUL prediction. However, these methods involve shallow architectures, which are less flexible in modeling the relationship between monitoring data and RUL [18]. Deep learning, as an advanced branch of machine learning, models complicated processes via hierarchical representations and has been investigated for system prognosis [19]. Chemali *et al.* [20] developed an RNN with LSTM for performing accurate state of charge estimation of lithium-ion batteries. Huang *et al.* [21] proposed a deep coupling convolutional neural network for intelligent compound fault diagnosis. The gearbox fault tests were used to verify its effectiveness. Zhu *et al.* [22] presented a deep feature learning method for RUL prediction through time–frequency representation and multiscale convolutional neural network. The effectiveness was validated by real experiment data. Miao *et al.* [23] developed dual-task deep LSTM networks for degradation assessment and RUL prediction of aero-engines. Its effectiveness was validated using the public C-MAPSS lifetime datasets. Although many deep learning methods have been investigated in fault diagnosis and prognosis, the sequential and time-domain features are not considered simultaneously when performing prognosis with monitoring data.

III. Proposed Deep CLSTM Model

In this section, a problem statement of RUL prediction is formulated, and a theoretical background of LSTM is introduced. To address the limitations of the traditional network, the novel architecture called CLSTM is constructed, and finally, the details of deep architecture based on CLSTM and RUL prediction procedures are presented.

A. Problem Statement

Vibration-based techniques have proved to be one of the most effective ways for bearing fault diagnosis among various techniques and are widely used to monitor the degradation process of bearings [24]. In the process of bearing's degradation, the vibration signals collected from sensors exhibit nonstationary characteristic; thus, the investigation of time–frequency analysis is important and necessary for RUL prediction. The time–frequency analysis provides the information in both time and frequency domains, which can reflect the change in frequency over the bearing's degradation. The result of time–frequency analysis for vibration signals is a 2-D matrix, which is used for RUL prediction rather than vibration signals in the time domain. This problem can be regarded as a prediction problem with a sequence of spatiotemporal features in both time and frequency domains.

Since vibration signals collected from bearings' run-to-failure tests exhibit a nonstationary characteristic, only time-domain

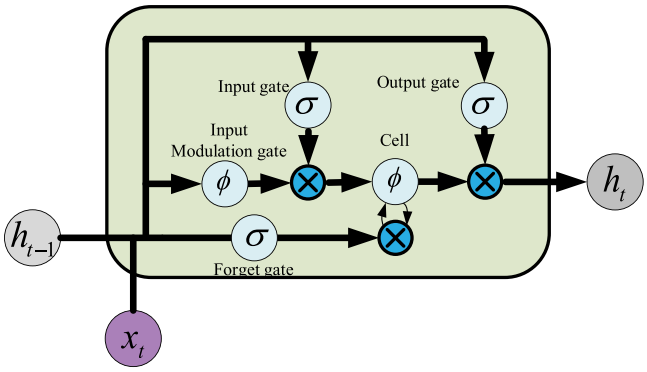


Fig. 1. Structure of the LSTM network [27].

or frequency-domain analysis fails to provide valuable degradation information. Thus, time–frequency analysis has been employed for nonstationary vibration signals, which can provide the information both when and at what frequencies a signal event happens compared with statistical features. The short-time Fourier transform (STFT) is widely used in rotation machinery condition monitoring because it gives information both in time and frequency domains. The STFT computes the Fourier transform of vibration signals $v(t)$ over a real and symmetric window function $\eta(t)$ [25]. The definition of the STFT is expressed as

$$S(\tau,\eta)=\int_{-\infty}^{+\infty} v(t)\eta(t-\tau)\,e^{-j\eta t}dt \tag{1}$$

where η and τ indicate modulated frequency and translated time, respectively. S is a function of both frequency η and time τ. When a sliding window goes over the time-domain signals, the spectral components in the block are calculated as function of frequency.

Suppose the number of datasets for bearings' vibration signals is T; then, the time–frequency of each dataset can be represented by a matrix $S_i\ (i=1,2,...,T)$. Bearing's degradation is a gradual process over time; thus, when predicting the current RUL with monitoring data, previous signals should be considered because the current RUL is related to the previous state. That is to say, RUL prediction is similar to the time-series-based problem. The formula of RUL $\widehat{r}_t$ prediction of time t based on the sequence matrix $S_{t-l+1}, S_{t-l+2}, ..., S_t$ is defined as

$$\widehat{r}_t=\arg\max P\left(\widehat{r}_t|S_{t-l+1},S_{t-l+2},...,S_t\right). \tag{2}$$

B. Long Short-Term Memory Network

LSTM is a special RNN structure that has recently become a popular choice for modeling temporal sequences and long-range dependencies in various tasks such as language modeling, learning word embedding, online handwritten recognition, and speech recognition [26]. The main advantage of LSTM lies in overcoming some shortcomings of RNNs, such as the vanishing and exploding gradients in the training process. The basic structure of LSTM is shown in Fig. 1, which is composed of one input layer, one recurrent layer, and one output layer. The core parts behind LSTM are the invention of gates, including hidden unit h_t, input gate i_t, forget gate f_t, output gate o_t, input modulation gate g_t, and memory cell c_t. The information flowing to the memory cell from the input activations is determined by the input gate. Meanwhile, the output gate conditionally controls what to output to the rest of the network based on the output activations. The forget gate controls how much information is thrown away of the internal state of the cell before it imports to the cell through the self-recurrent connection, which aims to forget or reset the cell's memory.

Memory cell and gates are used to decide what information should be kept and forgot, which leads the gradient to be trapped in the cell and prevented from vanishing too quickly. When the input gate i_t is turning ON, the memory cell c_t allows the information to be added to the cell. In this process, if the forget gate f_t is activated, the information from previous cell state c_{t-1} could be ignored. The output gate o_t will further determine the information flow from the updated cell output c_t to be propagated to the final state h_t. The input, cell output, and states are all 1-D vectors. To learn the precise timing of the outputs, peephole connections from its internal to the gates in the same cell are introduced. The update process of LSTM for time step t given inputs x_t, h_{t-1}, and c_{t-1} is shown as follows:

$$i_t=\sigma\left(W_{xi}x_t+W_{hi}h_{t-1}+W_{ci}\odot c_{t-1}+b_i\right) \tag{3}$$

$$f_t=\sigma\left(W_{xf}x_t+W_{hf}h_{t-1}+W_{cf}\odot c_{t-1}+b_f\right) \tag{4}$$

$$g_t=\phi\left(W_{xc}x_t+W_{hc}h_{t-1}+b_c\right) \tag{5}$$

$$c_t=f_t\odot c_{t-1}+i_t\odot g_t \tag{6}$$

$$o_t=\sigma\left(W_{xo}x_t+W_{ho}h_{t-1}+W_{co}\odot c_t+b_0\right) \tag{7}$$

$$h_t=o_t\odot\phi\left(c_t\right) \tag{8}$$

where W terms are corresponding weight matrices (e.g., W_{xi} denotes the matrix of weights from the input gate to the input). W_{ci}, W_{cf}, and W_{co} denote diagonal weight matrices for peephole connections, the b terms are the bias vectors (b_i denotes the input gate bias vector), $\odot$ is the elementwise product of the vectors, $\sigma(x)$ is used to squash the inputs to a [0, 1] range, which is defined as the sigmoid nonlinearity, and $\phi(x)$ is the hyperbolic tangent nonlinearity, similarly squashing its inputs to a $[-1, 1]$ range, as shown in the following:

$$\sigma(x)=\frac{1}{1+e^{-x}} \tag{9}$$

$$\phi(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}. \tag{10}$$

As a multiple-layer and hierarchical architecture is more efficient and powerful in feature representation than a shallow model, the same benefits can be harnessed with LSTMs, which means additional layers can be added to extract abstracted features over time for sequential data. The sequence output of one-layer LSTM is used as the input of another LSTM, which are widely used to deal with the challenge of sequence prediction problems.

C. Architecture of CLSTM

Learning representations of vibration time–frequency features with convolutional operation have shown clear advantages

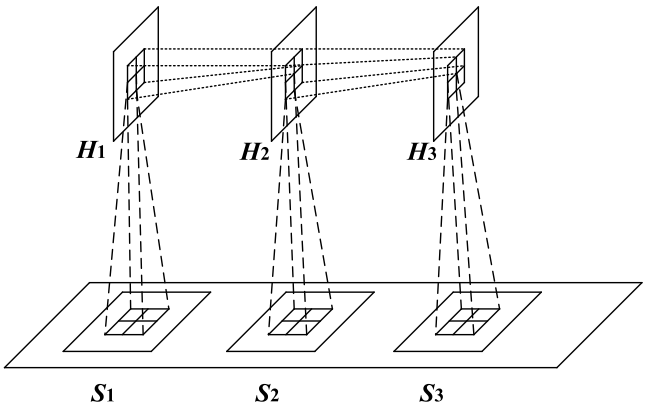


Fig. 2. Structure of CLSTM.

over "hand-crafted" features in bearing's diagnosis and prognosis [28], [29]. Extensions of CNN representations to model sequence have been proposed in recent works [30]. When statistical features of vibration signals are used as inputs of LSTM for RUL prediction, the time–frequency information is ignored. To address this problem, the CLSTM structure is constructed, whose inputs are time–frequency features, namely 2-D matrices $S_1, S_2, \ldots$. What is different from the traditional LSTM is that the proposed model introduces convolutional operation in both input-to-state and state-to-state transitions, as shown in Fig. 2. Convolutional operation can learn the meaningful features automatically given the time-domain features [31]. The benefit of the CLTMS is to model the sequences of time–frequency features instead of statistical features. We extend feature representation with CNN to significantly longer temporal convolutions, which enables RUL prediction at their long temporal scale. Similar to standard LSTM, CLSTM also consists of input gates, output gates, forget gates, cells, and output gates. For the implementation of CLSTM, it mainly includes unrolling the forward pass and backward pass. The equations for calculating the outputs in the forward process are shown as follows:

input gates:

$$A_i^t = W_{si} * S_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + B_i \tag{11}$$

$$Y_i^t = \sigma\left(A_i^t\right) \tag{12}$$

forget gates:

$$A_f^t = W_{sf} * S_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + B_f \tag{13}$$

$$Y_f^t = \sigma\left(A_f^t\right) \tag{14}$$

cells:

$$A_c^t = W_{sc} * S_t + W_{hc} * H_{t-1} + B_c \tag{15}$$

$$Y_c^t = Y_f^t \odot Y_c^{t-1} + Y_i^t \odot \phi\left(A_c^t\right) \tag{16}$$

output gates:

$$A_o^t = W_{so} * S_t + W_{ho} * H_{t-1} + W_{co} \odot Y_c^t + B_o \tag{17}$$

$$Y_o^t = \sigma\left(A_o^t\right) \tag{18}$$

Cell outputs:

$$H_o^t = Y_o^t \odot \phi\left(Y_c^t\right) \tag{19}$$

where $*$ denotes the convolution operator, and W and B terms denote the weights and biases needed to learn. $(H_o^t)_{\text{flatten}}$ means a matrix is flattened into a vector.

When it is used to make predictions, a logistic layer is added on the cell outputs, shown as follows:

$$\widehat{r}_t = \sigma(W_{hr}\left(H_o^t\right)_{\text{flatten}} + b_r). \tag{20}$$

Assuming that the actual RUL at time t is r_t, the loss function of the CLSTM is defined as

$$L\left(S, \theta\right) = \min \frac{1}{2}\|r_t - \widehat{r}_t\|^2 \tag{21}$$

where θ represents the model's parameters: $\theta = \{W_{si}, W_{hi}, W_{ci}, W_{sf}, W_{hf}, W_{cf}, W_{sc}, W_{hc}, W_{so}, W_{ho}, W_{co}, W_{hr}, B_i, B_f, B_c, B_o, b_r\}$. The objective function is a nonconvex problem with huge input data. Thus, stochastic gradient descent (SGD) is utilized to solve the optimization. In the process, as with standard LSTM, backpropagation through time is used to compute the gradient with respect to the cost in order to optimize the parameters. For the top logistic layer of prediction, the gradients of loss function with respect to the weight and bias can be calculated as

$$\frac{\partial L}{\partial W_{hr}} = -\left(r_t - \widehat{r}_t\right)\sigma'\left(\cdot\right)\left(H_o^t\right)_{\text{flatten}} \tag{22}$$

$$\frac{\partial L}{\partial b_r} = -\left(r_t - \widehat{r}_t\right)\sigma'\left(\cdot\right). \tag{23}$$

The calculation of other parameters' gradients is similar to BPTT in LSTM because CLSTM only contains the convolution operation. The difference is the gradients of the kernel weights and bias, which need to add the overall

$$\frac{\partial L}{\partial W_{co}} = \sum_{u,v}\left(\delta\right)_{uv}\left(p\right)_{uv} \tag{24}$$

where $(p)_{uv}$ is the patch that is multiplied elementwise by W_{co} during the convolution in order to compute the element at (u, v) in the output convolution map. After the gradients are calculated with backpropagation, the parameters can be optimized with SGD. The update process of the parameters can be written as

$$\theta = \theta - \alpha\frac{\partial L}{\partial \theta} \tag{25}$$

where α is the learning rate.

The inputs, cell outputs, and hidden states of the CLSTM can be seen as an extension of traditional LSTM represented by (6) with the two dimensions. The transitional kernels of CLSTM can capture the dynamic properties. The size of convolutional output is kept the same number by padding in the dimensions. Zero-padding operation of hidden states on the boundary points is implemented such that the obtained output has the same dimension as the original input. In the initialization, all the states of the LSTM are set to 0, which means "total ignorance" of the future. Similarly, zero padding of hidden states corresponds to no prior knowledge about the initial states.

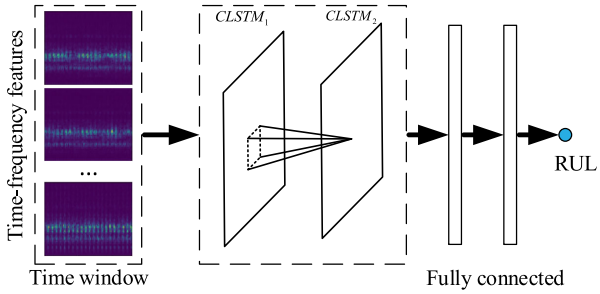


Fig. 3. Deep CLSTM for RUL prediction.

D. Deep CLSTM Model for RUL Prediction

Like traditional LSTM, CLSTM can be employed as a building block for deep architecture, which has been proven to be highly effective in various application due to its ability to learn hierarchical representation of the raw data [32]. Although LSTM has deep architecture in the view of it is considered as a feedforward neural network unrolled in time where each layer shares the same model parameters, the features are only learned by a single nonlinear layer at a given time before contributing the output for that time instant. Thus, stacking multiple layers allows us to learn at different time scales over the input, which can better use parameters by distributing them over the space through multiple layers.

The overall architecture of the proposed method for RUL is presented in Fig. 3. The inputs are time–frequency features of vibration signals, which are followed by the CLSTM layer aimed to model sequences of time–frequency features. Then, the 2-D feature map is flattened and fully connected with a rectified linear unit layer. At the end, the RUL prediction is provided through a linear regression layer, as the output of the deep neural network. The parameters of the proposed model are updated by minimizing the training error. The Adam algorithm [33] is employed in the process of optimization. The construction of deep architecture is expected to have strong representation power, which aims to learn the high-level representations for time–frequency features. The fully connected layer uses all the learned features for final regression to predict RUL. The proposed method can learn the spatial relationships among time–frequency features as well as intrinsic temporal dependence.

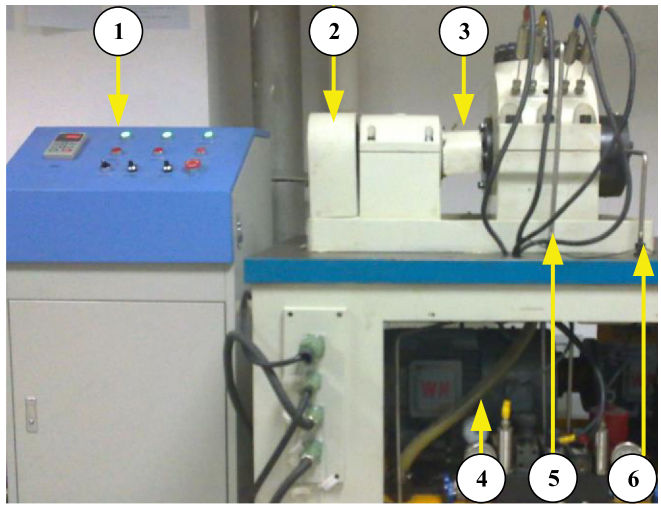
IV. Application and Validation

To validate the effectiveness of the proposed method, bearings' run-to-failure tests were carried out on test rig. Vibration signals are collected in the experiment. Then, RUL of bearings are performed using the proposed method.

A. Experiment

In order to validate the effectiveness of the proposed method and truly reflect the real defect propagation process, bearing run-to-failure tests were carried out under constant load conditions. The test rig consisted of a power and drive system, the main body, a hydraulic loading system, a lubrication system, a control system, and an independent data collection system, as shown in Fig. 4. The bearing test rig host four test bearings on one shaft, which is designed as a supported beam structure. Two test bearings were installed on both ends of the shaft, while two steady bearings, N312 cylindrical roller bearings, were fixed at the middle of the shaft, as shown in Fig. 5. The axial load was inflicted on bearing 4 and transmitted to bearing 1 by the shaft. The radial load is added to steady bearings directly and inflicted on test bearings through the shaft. The parameters of test bearings and steady bearings are presented in Table I.

In the experiment, the rotation speed was kept constant at 2500 r/min. The radial load was set to be 12 kN. Four thermocouple temperature sensors were employed to monitor the temperature of test bearings and steady bearings, respectively. The test would stop when the temperature exceeds a certain level, which indicates the failure of test bearings. Vibration signals, collected with



1: Control cabinet, 2: Belt transmission, 3: Flexible coupling, 4: Motor, 5: Racial load, 6: Axial load

Fig. 4. Bearing test rig of the experiments.

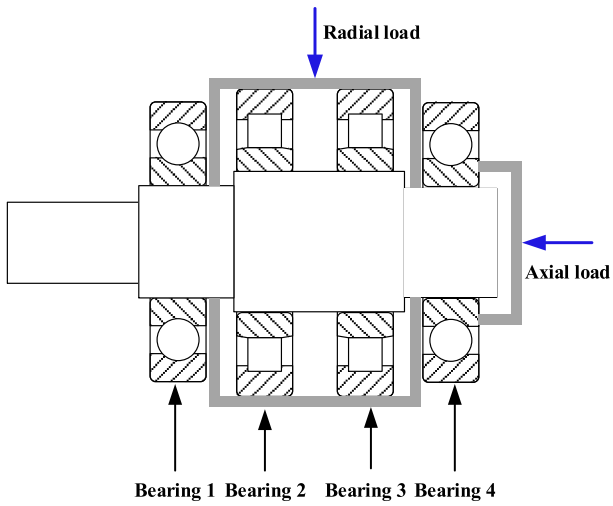


Fig. 5. Structure of bearing loading.

TABLE I
PARAMETERS OF TEST AND SUPPORT BEARINGS

Bearing type	Inner diameter (mm)	Outer diameter (mm)	Roller diameter (mm)	Roller Number
60911	55	80	7.1	16
N312	60	130	19.1	16

(a) (b)

Fig. 6. Defects of experimental bearings. (a) Bearing A. (b) Bearing B.

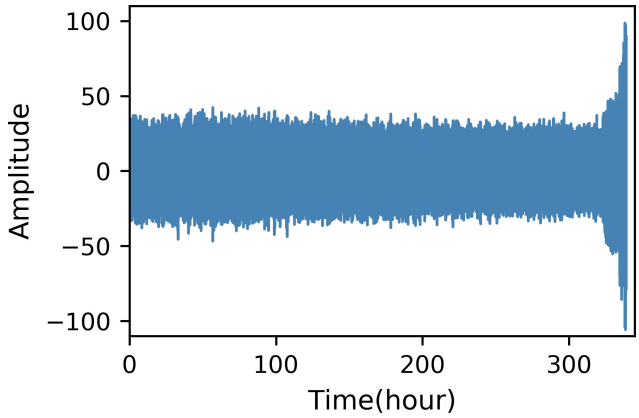


Fig. 7. Run-to-failure vibration signals of bearing A.

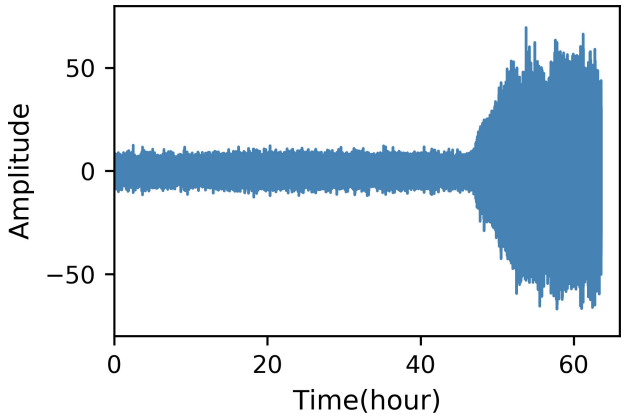


Fig. 8. Run-to-failure vibration signals of bearing B.

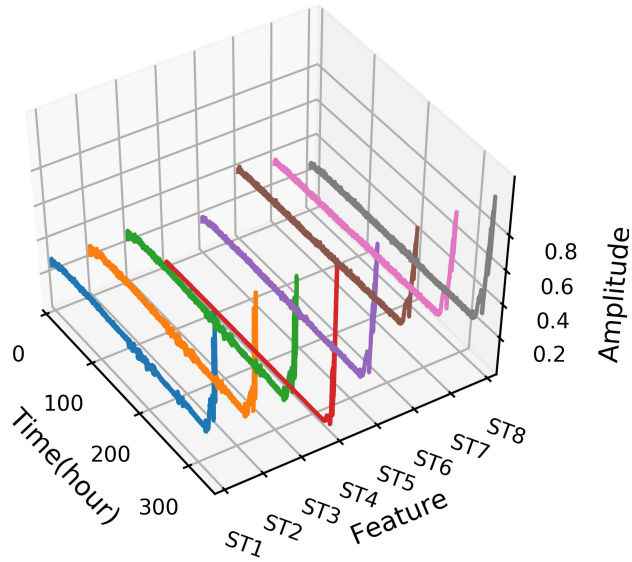


Fig. 9. Statistical features of bearing A.

a sampling rate of 20 480 Hz, were transmitted by the screw that contacted with bearing's outer race. Each sample set contained 32 768 points and was collected every 5 min.

B. RUL Prediction

Two sets of tests were carried out successfully. Test 1 ended with various defects in bearing 4, namely inner race, outer race, and rolling element defects. Test 2 ended up with an inner race defect in bearing 4. The inspections photos of from tests 1 and 2 are shown in Fig. 6. In the experiments, vibration signals of bearing A (bearing 1 in test 1) and bearing B (bearing 1 in test 2) for the whole lifetime are shown in Fig. 7 and 8. There are 4071 and 763 data sets for bearings A and B, respectively. The statistical features in time and frequency domains, such as root mean square, square root value, absolute mean, kurtosis, and so on, are able to describe the degradation process of bearings. They are depicted in Fig. 9 and 10 over the entire life cycles

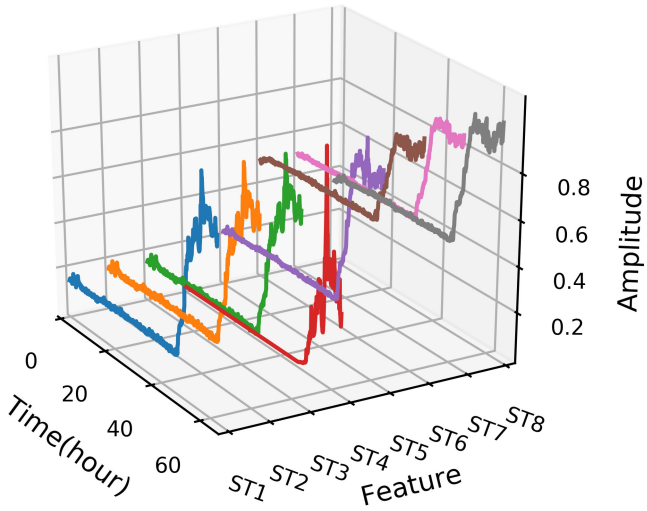


Fig. 10. Statistical features of bearing B.

TABLE II
STATISTICAL FEATURES OF VIBRATION SIGNALS

Time-domain features	Formula
ST_1	$\sqrt{\frac{1}{n}\sum_{i=1}^{n} v_i^2}$
ST_2	$\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\lvert v_i\rvert}\right)^2$
ST_3	$\frac{1}{n}\sum_{i=1}^{n}\lvert v_i\rvert$
ST_4	$\frac{1}{n}\sum_{i=1}^{n}\left(v_i-\frac{1}{n}\sum_{i=1}^{n} v_i\right)^4$
Frequency-domain features	**Formula**
ST_5	$\frac{1}{N}\sum_{j=1}^{N}\Omega_j$
ST_6	$\sqrt{\frac{\sum_{j=1}^{N}(\omega_j-ST)^2\Omega_j}{N}}$
ST_7	$\frac{ST_6}{ST},\ ST=\frac{\sum_{j=1}^{N}\omega_j\Omega_j}{\sum_{j=1}^{N}\omega_j}$
ST_8	$\frac{\sum_{j=1}^{N}\sqrt{\lvert\omega_j-ST\rvert}\Omega_j}{N\sqrt{ST_6}}$

where v_i and n denote the vibration signals and the number of points, respectively. Ω_j and ω_j denote the frequency spectrum and frequency components, respectively. N is the length of frequency series.

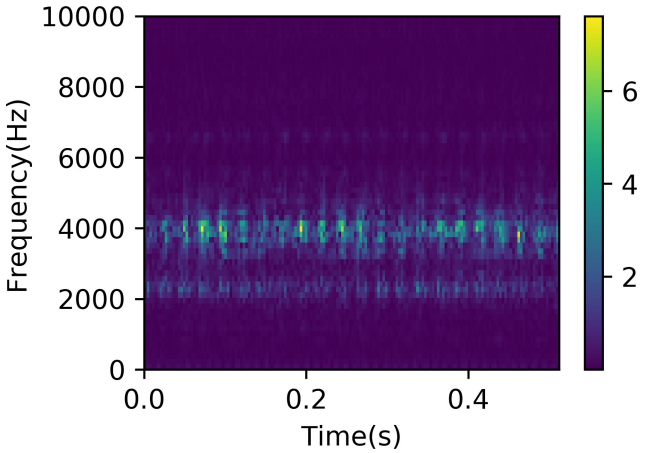


Fig. 11. Time–frequency feature at 41.7 h.

of bearings A and B, respectively. The definition of statistical features is presented in Table II. It is observed the most of bearings' life is spent on the material accumulative stage, while the time from occurrence of initial defect to the final failure is relatively short. When the initial fault occurred, the damage grows rapidly, especially near the final failure. Therefore, RUL prediction before failure occurs can provide enough time for maintenance and logistical scheduling.

In the time-series-based problem such as RUL prediction, more information can be generally obtained from temporal signals compared with signals sampled at a single time step. Time sequence techniques own a large potential for better prediction performance, since they consider temporal information. Time–frequency features contain more information; the time–frequency features at time 41.7 and 333.3 h of bearing A are shown in Figs. 11 and 12. At each time step, the past time–frequency features within the time window $l = 6$ were employed as inputs of the network.

For bearings A and B, the last 100 and 50 data sets of bearing's life were used as the testing sets, respectively. For the training

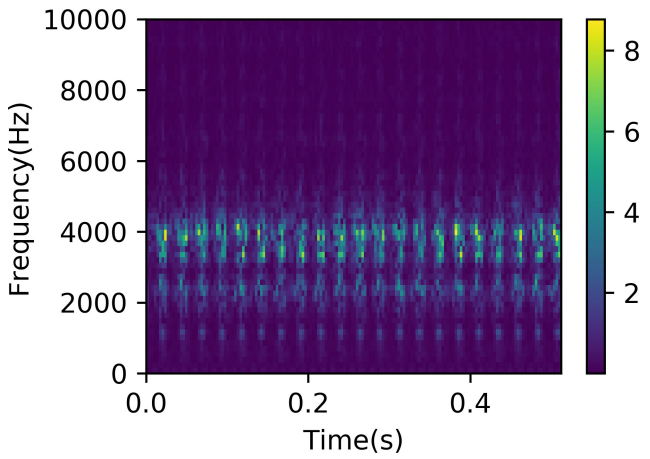


Fig. 12. Time–frequency features at 333.3 h.

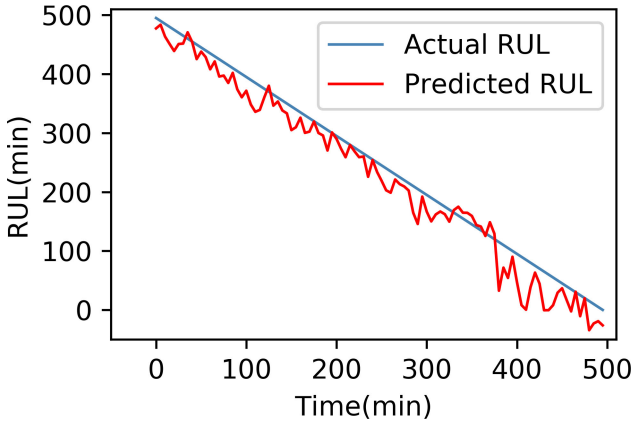


Fig. 13. RUL prediction of bearing A.

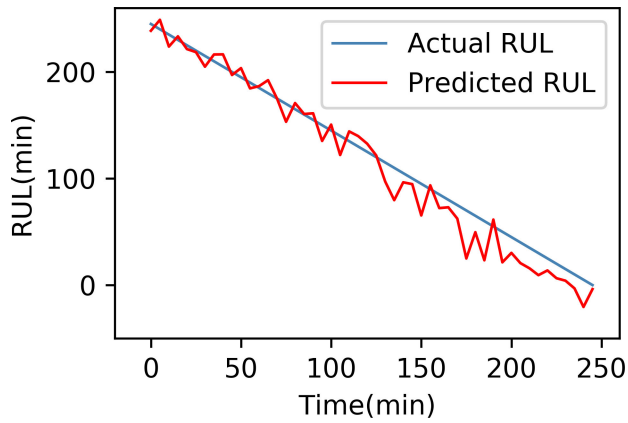


Fig. 14. RUL prediction of bearing B.

set, the last 301 and 207 data sets (except the testing sets) were used. The vibration signals of each data set were transformed into time–frequency features by the STFT, which were used as inputs of deep CLSTM. The prediction results are shown in Fig. 13 and 14, respectively. To simplify the parameter selection, in each layer, the number of the hidden units and the size of

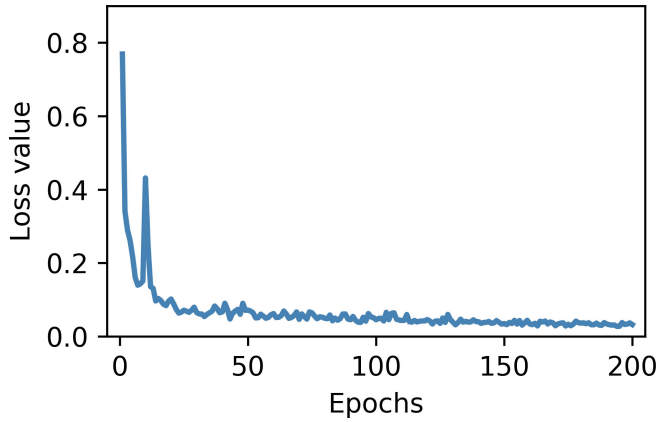


Fig. 15. Loss value during the training process.

convolutional kernel are set the same. The grid search is adopted in the range of [60, 120] to find the optimal value of the number of hidden units, which is 110. The batch size is 10 considering that there are 200 training samples totally. The input size of the model is $6 \times 165 \times 41$. The hidden layer of CLSTM is 2 and the convolutional filter is 5×4, For the deep architecture with two-layer CLSTM, the Adam optimizer is employed in the optimization. The rectified linear unit activation function is used as it empirically provides good results and avoids the vanishing gradient problem that other nonlinear activation functions can cause. The learning rate was 0.001. The loss function for bearing A in the training process can be seen in Fig. 15. After 50 epochs, the training error almost converges to 0.

V. Results and Discussion

To show the superiority of the CLSTM model, the RUL prediction performance of the proposed method is compared with those of other network architectures, including deep LSTM model and deep CNN model. The deep LSTM model takes the statistical features as inputs instead of time–frequency features. The statistical features of time and frequency domains are defined in Table II. The time window is set the same as that of the deep CLSTM model. The deep CNN model relies on the time–frequency features without previous information, which means that only current features are used to predict the RUL directly without considering the vibration signals of the previous step. To evaluate the prediction results, two metrics, namely mean absolute error (MAE) and root mean squared error (RMSE), which defined by (26) and (27), respectively, are used to calculated the errors:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|r_i - \widehat{r}_i| \tag{26}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(r_i - \widehat{r}_i)^2} \tag{27}$$

where r_i and $\widehat{r}_i$ are actual and predicted RULs, respectively.

The results of RUL prediction using different methods are presented in Table III and Fig. 16 and 17. Accurate RUL predictions

TABLE III
RUL Prediction Results

	Model	Deep CLSTM	Deep CNN	Deep LSTM
Bearing A	RMSE (min)	29.7	47.9	40.7
	MAE (min)	23.5	40.1	35.2
Bearing B	RMSE (min)	10.8	18.2	15.6
	MAE (min)	6.0	12.6	9.7

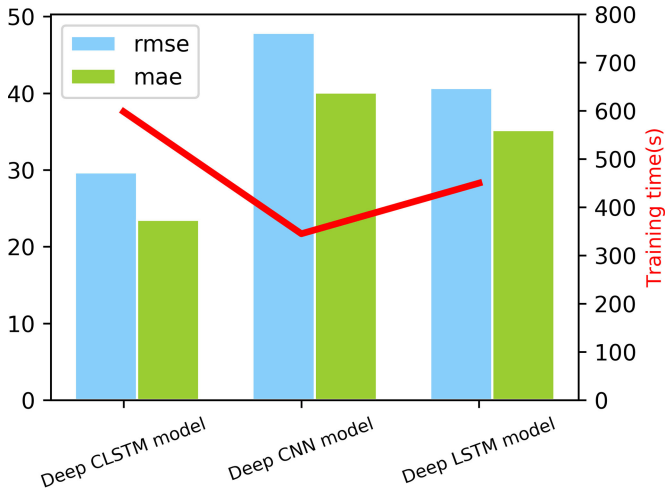


Fig. 16. RUL prediction by different methods for bearing A.

Fig. 17. RUL prediction by different methods for bearing B.

based on the proposed method indicate that the CLSTM model can capture long-term dependencies of time–frequency features of vibration signals more efficiently, whereas the deep CNN model without previous information and the deep LSTM model without time–frequency features fail to learn long-term spatial features. This is because time–frequency features contain more information as compared with statistical features. However, the training time of the deep CLSTM model is longer than that of other methods because the convolution operation costs much time.

VI. Conclusion

In this article, a novel deep learning algorithm was proposed for ball bearing RUL prediction, which was critical to PHM of a variety of rotating machineries. Specifically, a new extension of LSTM network with convolutional operation integrated, named CLSTM, was proposed in this article. From the theoretical standpoint, CLSTM preserved the advantage of conventional LSTM, which maintained the temporal dependencies and selectively propagated relevant information through the system dynamics. In addition, the proposed CLSTM also extracted the time–frequency features given its inherent convolutional characteristics. By stacking CLSTMs, a deep architecture for RUL prediction was constructed to extract high-level information. Run-to-failure tests were conducted on rotating machines to provide a benchmark to validate the proposed framework. In the experiments, vibration responses of ball bearings were collected and adopted for network training and RUL prediction. The performance of the proposed CLSTM was proved to be much enhanced by comparing to the state of the art, despite a slight increase of training time. Given the promising capability of CLSTM in integrating the temporal dependence with time–frequency characteristics, the future work of RUL prediction is suggested to be architecture optimization, in order to reduce the training time, as well as to further enhance the applicability for realistic operation scenarios. Moreover, the prediction approach will be applied and evaluated to more sophisticated platforms with multiple components.

References

[1] J. Lee, F. J. Wu, W. Y. Zhao, M. Ghaffari, L. X. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems-reviews, methodology and applications," *Mech. Syst. Signal Process.*, vol. 42, nos. 1/2, pp. 314–334, 2014.

[2] M. Ma, C. Sun, and X. Chen, "Deep coupling autoencoder for fault diagnosis with multimodal sensory data," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1137–1145, Mar. 2018.

[3] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.

[4] M. Ma, C. Sun, C. Zhang, and X. Chen, "Subspace-based MVE for performance degradation assessment of aero-engine bearings with multimodal features," *Mech. Syst. Signal Process.*, vol. 124, pp. 298–312, 2019.

[5] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, and X. Chen, "Deep transfer learning based on sparse auto-encoder for remaining useful life prediction of tool in manufacturing," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2416–2425, Apr. 2019.

[6] M. Ma, X. Chen, X. Zhang, B. Ding, and S. Wang, "Locally linear embedding on Grassmann manifold for performance degradation assessment of bearings," *IEEE Trans. Rel.*, vol. 66, no. 2, pp. 467–477, Jun. 2017.

[7] W. Peng, Z.-S. Ye, and N. Chen, "Joint online RUL prediction for multi-deteriorating systems," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2870–2878, May 2019.

[8] M. Daigle and K. Goebel, "Model-based prognostics under limited sensing," in *Proc. Aerosp. Conf.*, 2010, pp. 1–12.

[9] Y. Hu, P. Baraldi, F. Di Maio, and E. Zio, "Online performance assessment method for a model-based prognostic approach," *IEEE Trans. Rel.*, vol. 65, no. 2, pp. 718–735, Jun. 2016.

[10] A. El Mejdoubi, H. Chaoui, J. Sabor, and H. Gualous, "Remaining useful life prognosis of supercapacitors under temperature and voltage aging conditions," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4357–4367, May 2018.

[11] H. Sun, D. Cao, Z. Zhao, and X. Kang, "A hybrid approach to cutting tool remaining useful life prediction based on the wiener process," *IEEE Trans. Rel.*, vol. 67, no. 3, pp. 1294–1303, Sep. 2018.

[12] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, 2019.

[13] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.

[14] M. Ma, C. Sun, X. Chen, X. Zhang, and R. Yan, "A deep coupled network for health state assessment of cutting tools based on fusion of multisensory signals," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6415–6424, Dec. 2019.

[15] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional bi-directional LSTM networks," *Sensors*, vol. 17, no. 2, 2017, Art. no. 273.

[16] X. Chen, Z. Shen, Z. He, C. Sun, and Z. Liu, "Remaining life prognostics of rolling bearing based on relative features and multivariable support vector machine," *Proc. Ins. Mech. Eng., C: J. Mech. Eng. Sci.*, vol. 227, no. 12, pp. 2849–2860, 2013.

[17] M. Elforjani and S. Shanbr, "Prognosis of bearing acoustic emission signals using supervised machine learning," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5864–5871, Jul. 2018.

[18] M. Ma, C. Sun, and X. Chen, "Discriminative deep belief networks with ant colony optimization for health status assessment of machine," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 12, pp. 3115–3125, Dec. 2017.

[19] C. Sun, M. Ma, Z. Zhao, and X. Chen, "Sparse deep stacking network for fault diagnosis of motor," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3261–3270, Jul. 2018.

[20] E. Chemali, P. J. Kollmeyer, M. Preindl, R. Ahmed, and A. Emadi, "Long short-term memory networks for accurate state-of-charge estimation of li-ion batteries," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6730–6739, Aug. 2018.

[21] R. Huang, Y. Liao, S. Zhang, and W. Li, "Deep decoupling convolutional neural network for intelligent compound fault diagnosis," *IEEE Access*, vol. 7, pp. 1848–1858, 2018.

[22] J. Zhu, N. Chen, and W. Peng, "Estimation of bearing remaining useful life based on multiscale convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3208–3216, Apr. 2018.

[23] H. Miao, B. Li, C. Sun, and J. Liu, "Joint learning of degradation assessment and RUL prediction for aero-engines via dual-task deep lstm networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5023–5032, Sep. 2019.

[24] S. Wang, I. Selesnick, G. Cai, Y. Feng, X. Sui, and X. Chen, "Nonconvex sparse regularization and convex optimization for bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 65, no. 9, pp. 7332–7342, Sep. 2018.

[25] S. Wang, X. Chen, I. W. Selesnick, Y. Guo, C. Tong, and X. Zhang, "Matching synchrosqueezing transform: A useful tool for characterizing signals with fast varying instantaneous frequency and application to machine fault diagnosis," *Mech. Syst. Signal Process.*, vol. 100, pp. 242–288, 2018.

[26] J. Donahue *et al.*, "Long-term recurrent convolutional networks for visual recognition and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 2625–2634.

[27] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in *Proc.15th Annu. Conf. Int. Speech Commun. Assoc.*, 2014, pp. 338–342.

[28] W. Sun, R. Zhao, R. Yan, S. Shao, and X. Chen, "Convolutional discriminative feature learning for induction motor fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1350–1359, Jun. 2017.

[29] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727–2736, Mar. 2018.

[30] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

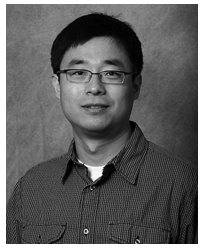
[31] Z. Chen, K. Gryllias, and W. Li, "Mechanical fault diagnosis using convolutional neural networks and extreme learning machine," *Mech. Syst. Signal Process.*, vol. 133, 2019, Art. no. 106272.

[32] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2015, pp. 4580–4584.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Rep. (ICLR)*, 2015.

Meng Ma received the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2018.

He is currently a Postdoctoral Research Associate with the Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, MA, USA. His current research interests include prognostics and health management, machine learning, and risk assessment.



Zhu Mao received the B.S. degree in automotive engineering from Tsinghua University, Beijing, China, in 2002, and the M.S. and Ph.D. degrees in structural engineering from the University of California San Diego, La Jolla, CA, USA, in 2008 and 2012, respectively.

He is currently with the Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, MA, USA. He has authored or coauthored more than 60 articles in journals and conference proceedings. He is the author of one book chapter. His current research interests include structural dynamics and health monitoring, noncontact sensing and image processing, and uncertainty quantification.

Dr. Mao is the recipient of the D. J. DeMichele Scholarship Award from the Society for Experimental Mechanics (SEM) in 2011, the Young Investigator Award from the U.S. Air Force Office of Scientific Research in 2018, and the SAGE Publishing Young Engineer Lecture Award from the SEM in 2019. He is the member of the American Society of Mechanical Engineers, the SEM, and the International Society for Optics and Photonics.