SPECIAL SECTION ON DATA ANALYTICS AND ARTIFICIAL INTELLIGENCEFOR PROGNOSTICS
AND HEALTH MANAGEMENT (PHM) USING DISPARATE DATA STREAMS

IEEE *Access*
Multidisciplinary : Rapid Review : Open Access Journal

Cascade Fusion Convolutional Long-Short Time Memory Network for Remaining Useful Life Prediction of Rolling Bearing

QIONG WU [ID]1 AND CHANGSHENG ZHANG [ID]2
[1]School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China
[2]Software College, Northeastern University, Shenyang 110169, China

Corresponding author: Changsheng Zhang (zhangchangsheng@mail.neu.edu.cn)

ABSTRACT Deep learning has been a widely adopted approach to achieve the remaining useful life prediction (RUL) of rolling bearing. However, the architectures of the current proposed deep learning approaches are limited and the prediction result is less stable on account of the single sensory data adopted. To address this issue, a new cascade fusons cascade convolutional long-short time memory network is proposed for bearing RUL prediction, in which a cross connection block is formulated to fuse the information streams from the adjacent channels twice and a concentration operation is also affiliated in the end of the network to integrate the separated information streams into an ensemble form. Meanwhile, a convolutional long-short time memory network is adopted as the basic cell in the proposed network on account of its ability to reflect the spatial-temporal correlation of the representative features. Moreover, a smoothing method based on multi-averaging operation is constructed in the prediction phase to largely eliminate the fluctuation in the prediction results. The application on the experimental bearing degradation dataset is able to verify the superiority and stability of the proposed method in comparison with the other comparison methods.

INDEX TERMS RUL prediction, bearing, deep learning, convolutional long-short time memory network.

I. INTRODUCTION

Since rolling bearing has been extensively used as the main supporting component in rotating machines, its performance will greatly affect the reliability and stability of the whole mechanical system. During the operating process of the rolling bearing, it is inevitable that its performance will be degraded with the increase of operation times [1]. As the degradation degree increases, it will further lead to a shutdown of the whole operating system, resulting in huge economic loss and serious safety problem [2], [3]. A remaining useful life (RUL) prediction technique for rolling bearing is urgently needed to effectively evaluate its healthy condition [4].

In recent years, a wide range of bearing RUL prediction techniques has been provided in reference and these techniques can be mainly divided into two groups [5]–[11]: one is the mechanism model based prediction method and the other is the data-driven based prediction method. The mechanism model can well reflect the relationship between the

The associate editor coordinating the review of this manuscript and approving it for publication was Faisal Khan [ID].

external measurable variables and the internal physical properties [5], [6]. However, the inherent mechanism knowledge of the bearing is needed in establishing the mechanism model and the established model is only suitable for the bearing with a specific condition [5], [6]. The data-driven based prediction method can make a full use of the monitoring data during the bearing degradation process to accurately predict its RUL. Therefore, a number of data-driven based prediction methods have been proposed in reference and most of them is established on the frequently used support vector machine, artificial neural network that owns only a shallow architecture contain a number of hidden layers or even without a hidden layer [7]–[10]. Though the prediction results was satisfied in these literatures, the manual involvement in the feature extraction process makes less automatics and the effective bearing RUL prediction method with high automatic level is urgently needed. With the development of artificial intelligence, deep learning has recently drawn significant attention in the RUL prediction of mechanical components. The recent advance in deep learning technique demonstrates that it owns a great capacity to process the monitoring data with high dimensional. A number of effective deep architectures

have been proposed in reference and these can be mainly concluded as: deep belief network (DBN) [11], [12], auto-encoder network (AEN) [12], convolutional neural network (CNN) [14]–[16] and the recurrent neural network (RNN) [17]–[19].

The data driven based prognostic techniques mentioned above were helpful to achieve the bearing RUL prediction. However, there still exist a number of imperfections in those prognostic techniques. For example, 1) the representative feature in the recorded signal is difficult to be fully exposed on account of the limitation of the network structure. 2) the utilization of single sensory data is easy to lead to an inaccurate and unstable prognostic result.

Since the multiple monitoring data from distinctive sensory data acquisitions is sufficient to capture the complete status information and the multiple information streams from various feature extraction operations are also helpful to yield a better and more accurate prognostic result, some researchers also tried to integrate multiple monitoring information streams into the deep learning based techniques to extract and fuse the significant features from multiple data acquisitions or multiple feature extraction operations. For example, Wang *et al.* [20] once proposed a deep separable convolutional network to achieve the prediction task, in which the monitoring data from two accelerometers were fused in the network and a separable volitional building was built to realize the representative feature extraction. Wu *et al.* [21] proposed a deep LSTM network to predict the RUL of mechanical component, in which the monitoring data from multiple sensors were fused to enhance the performance of the prediction result. The prognostic results in the above mentioned references illustrated that information fusion approach is effective to improve the accuracy and stability of the prognostic results in comparison with these techniques using only single sensory data. This also motivates us to establish an information fusion based network structure in this study.

Nevertheless, the other major concern in the existing techniques is that the architecture of the current proposed network only owns the capacity to depict the representative features among the spatial scale (the typical forms are DBN, AEN and CNN) or the ability to characterize the representative features among the temporal scale (the typical forms are RNN and its improved versions). However, few network structures are qualified to describe the representative feature among both spatial scale and temporal scale. Convolutional LSTM (ConvLSTM) cell proposed by Shi *et al.* [22] seems an effective choice to address this issue, in which the convolution operation is integrated inside the LSTM cell to make a full consideration of the spatial-temporal correlation of the representative feature in the recorded signal. Thus, to make an improvement of the current information fusion based prognostic methods, a cascade fusion ConvLSTM network (CF-ConvLSTM) is proposed to achieve the bearing RUL prediction in this study. In the proposed model, two feature extraction channels are formulated on the basis of the ConvLSTM cell to extract the representative features of the

monitoring data among two different directions and a cross connection block is applied on the architecture of network twice to transmit the representative features from one channel to the other channel to achieve the information fusion task initially. Meanwhile, a concentration operation is adopted in the end of network to fuse the representative features from two channels into one ensemble form once again. The proposed model is able to fuse the representative features from the multiple sensory data to provide a more accuracy and stable prognostic result.

The main contributions of this study are highlighted as follows:

1) We establish a CF-ConvLSTM model to extract the representative features form multiple sensory data for bearing RUL prediction. The proposed model is able to provide an accurate and stable prognostic result.
2) The cross connection block is first proposed in this study, which can effectively realize the goal of feature fusion between two adjacent channels.
3) To the best of our knowledge, this study first leverages the ConvLSTM cell to achieve the bearing RUL prediction task.
4) To eliminate the fluctuation of the prediction result, a multi-averaging method is also proposed in this study to largely smooth the prediction result.

The rest of this paper is arranged as follows. A brief review of the related works is given in Section II. Section III provides the theoretical background of the principle of the ConvLSTM. Section IV depicts the details of the architecture of the proposed network. Section V demonstrates the effectiveness of the presented RUL prediction method using an experimental bearing degradation dataset. Final, the conclusions are drawn in section VI.

II. RELATED WORK

The data-driven RUL prediction approaches in reference can be mainly summarized as two forms, i.e., the shallow architecture based RUL prediction method and the deep learning based RUL prediction method. The shallow architecture based RUL prediction method is mainly composed of two essential parts, i.e., feature extraction and regression prediction. The quality of the extracted feature is of vital importance for the performance of the prediction result. Ali *et al.* [8] tried to smooth the extracted features using the Weibull failure rate function and made an estimation of the degradation state via a specific neural network. Hong *et al.* [10] proposed a health trend prediction method for bearing, in which a confidence value that capable to reflect the bearing healthy condition was formulated on the basis of self-organizing map and several neural networks were adopted to predict the health trend in divided bearing healthy stages. Meanwhile, a reasonable regression prediction method is also helpful to improve the accuracy of the prediction result. Javed *et al.* [7] proposed an enhanced multivariate degradation model for RUL prediction of mechanical component, in which an improved maximum entropy fuzzy clustering algorithm was used in the first stage

to divide the degradation data into several stages and then an improved extreme learning machine was adopted to conduct the prediction task. Qu *et al.* [9] proposed a machine prognostic method by integrating the parameter optimized least square support vector regression with the cumulative sum trigger updating mechanism.

With the development of artificial intelligence, a number of effective deep architectures have been proposed in reference. The utilization of the deep architecture makes it unnecessary to formulate an additional feature extraction procedure in these RUL prediction methods on account of the deep architecture is able to effectively extract the significant features from the raw monitoring data. The architecture of DBN is stacked of a number of restricted Boltzmann machines layer by layer. Deutsch *et al.* [11] once proposed a deep belief network for RUL prediction of rotating components. Yu *et al.* [12] formulated a hybrid parameter optimization to optimize the key parameters of a constructed DBN model and further achieve the bearing fault mode classification. AEN is able to learn the representative features from the input data in an unsupervised approach and the extracted feature can be further used to achieve the classification or regression tasks. Ren et al. [13] tried to employ an AEN to compress the representative features from triple time domain, frequency domain and time-frequency domain, and then a deep neural network was proposed to achieve the bearing RUL prediction. Inspired by the researches of human brain visual cortex, the architecture of CNN was established in reference, which consists of a series of convolutional layers and pooling layers. The unique architecture of CNN makes it able to extract the meaningful information from two-dimensional image data. Wang *et al.* [14] tried to formulate a sample conversion method to make the recorded one-dimensional vibration data satisfy the input requirement of the two-dimensional CNN and further achieve the bearing RUL prediction. With the development of computer science, the appearance of one-dimensional CNN makes it more convenient to process the one-dimensional monitoring data. Thus, a number of RUL prediction methods based on one-dimensional CNN have been proposed. For example, Li *et al.* [15] also tried to achieve the bearing RUL prediction via the one-dimensional CNN. Yoo et al. [16] constructed a health indicator to reflect the bearing degradation trend using one-dimensional CNN, and then proposed a RUL prediction method on the variation of the HI in the initial period. To make a difference with the principle of the above mentioned three deep architectures, all nodes in RNN are linked in the chain and the parameters in RNN are recursively updated among the direction of sequence evolution, which can effectively depict the internal characteristic of time series data and is successfully used in the areas of natural language processing, speech recognition and handwriting recognition. However, a major problem in traditional RNN is that the gradient vanishing or exploding make it only suitable to capture the short-term memory. To address it, some improved versions of traditional RNN, i.e., long-short term memory (LSTM), gate recurrent

unit (GRU) and bidirectional LSTM, appear to construct different inner structures on the foundation of traditional RNN. Some RUL prediction works are also proposed on the basis of these improved RNN models. For example, Huang *et al.* [17] proposed a bidirectional LSTM network to simultaneously model both sensors and operational condition information in an integrated framework. Miao *et al.* [18] proposed a joint learning strategy on the foundation of LSTM to realize the bearing degradation assessment and RUL prediction in a dual output network. Chen *et al.* [19] proposed a GRU based recurrent neural network for bearing RUL prediction, in which the kernel principle component analysis was first used to achieve the feature reduction and the GRU network was then formulated to predict the RUL.

III. PRINCIPLE OF CONVLSTM

LSTM can be viewed as an improved version of the traditional RNN since it can effectively solve the problem of gradient vanishing or exploding in RNN. However, the full connections in the input-to-state and state-to-state transitions of the LSTM make it unavailable to encode the spatial information in the time series. ConvLSTM was formulated as a variant of LSTM by Shi et al. [22]. It owns a better performance to learn spatial information in the condition monitoring data on account of the convolution operation adopted in feature transitions. The main difference between LSTM and ConvLSTM is the number of the feature dimensions. The transmitted feature is one-dimensional in the LSTM while it can be represented as high dimensional in the ConvLSTM. The equations of the gates (input, forget and output) in ConvLSTM are as follows [22]:

$$\mathbf{I}_t = \sigma \left(\mathbf{W}_{xi} * \mathbf{X}_t + \mathbf{W}_{hi} * \mathbf{H}_{t-1} + \mathbf{W}_{ci} \circ \mathbf{C}_{t-1} + \mathbf{b}_i \right) \quad (1)$$

$$\mathbf{f}_t = \sigma \left(\mathbf{W}_{xf} * \mathbf{X}_t + \mathbf{W}_{hf} * \mathbf{H}_{t-1} + \mathbf{W}_{cf} \circ \mathbf{C}_{t-1} + \mathbf{b}_f \right) \quad (2)$$

$$\mathbf{o}_t = \sigma \left(\mathbf{W}_{xo} * \mathbf{X}_t + \mathbf{W}_{ho} * \mathbf{H}_{t-1} + \mathbf{W}_{co} \circ \mathbf{C}_{t-1} + \mathbf{b}_o \right) \quad (3)$$

$$\mathbf{C}_t = \mathbf{f}_t \circ \mathbf{C}_{t-1} + \tanh \left(\mathbf{W}_{xc} * \mathbf{X}_t + \mathbf{W}_{hc} * \mathbf{H}_{t-1} + \mathbf{b}_c \right) \quad (4)$$

$$\mathbf{H}_t = \mathbf{o}_t \circ \tanh \left(\mathbf{C}_t \right) \quad (5)$$

where ∘ means the Hadamard product, $\mathbf{I}_t$, $\mathbf{f}_t$, and $\mathbf{o}_t$ are input, forget and output gates. $\mathbf{W}$ represents the weight matrix, $\mathbf{X}_t$ denotes the current input data, $\mathbf{H}_{t-1}$ is previous hidden output, $\mathbf{C}_t$ and is the cell state. The convolution operation (*) is substituted for matrix multiplication between $\mathbf{W}$ and $\mathbf{X}_t$, $\mathbf{H}_{t-1}$ in the state-to-state and input-to-state transitions. Thus, the fully connection layer in LSTM can be replaced by the convolution layer in ConvLSTM and spatial information in the extracted feature is able to be well depicted.

IV. PROPOSED PROGNOSTIC MODEL FOR BEARING RUL PREDICTION

In this section, the principle of the new proposed cross connection block is first introduced in section 4.1 and then the details of the proposed CF-ConvLSTM network is described in section 4.2. After that, the principle of the training and testing processes for the proposed prognostic method is given in section 4.3. It is noted that since the input data in

this study is manifested as a one-dimensional time series, the one-dimensional convolutional layer is adopted in all convolutional related network structures in the proposed network.

A. CROSS CONNECTION BLOCK

The cross connection block is proposed in this study for the first time to realize the information stream fusion between two distant channels and its schematic is given in Fig. 1. The cross connection block contains two operations: information stream division and the information stream concentration. The principle of these two operations is illustrated as follows.
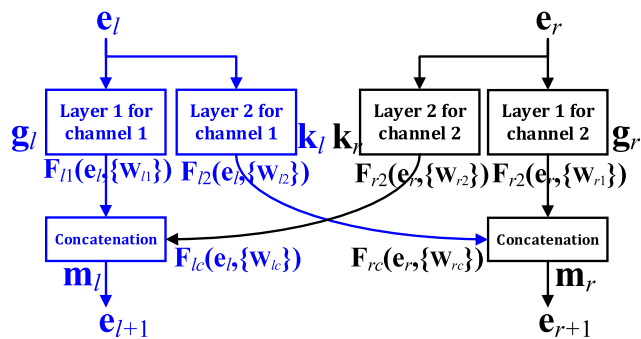

FIGURE 1. The schematic of the cross connection block.

The information stream from the left channel is selected to illustrate the principle of information stream division and the principle is also suitable for the information stream from the right channel. The original information stream is represented as with the dimension space of $m \times n \times k$. It can be separated into two streams $\mathbf{g}_l$ and $\mathbf{k}_l$ via the information stream division layers and the relationship between the separated streams and the original stream can be formulated as follows:

$$\mathbf{g}_l = \mathbf{F}_{l1}(\mathbf{e}_l, \{\mathbf{W}_{l1}\}) \qquad (6)$$

$$\mathbf{k}_l = \mathbf{F}_{l2}(\mathbf{e}_l, \{\mathbf{W}_{l2}\}) \qquad (7)$$

where $\mathbf{W}$ means the weight parameters in information stream division layers and $\mathbf{F}$ denotes the nonlinear function between the separated streams and the original stream. The dimension space of stream $\mathbf{g}_l$ is the same with that of the original stream $\mathbf{e}_l$, the dimension space of stream $\mathbf{k}_l$ is represented as $m \times n \times \omega$ and ω is less than k after the nonlinear function conducted.

In the information stream concentration, the stream $\mathbf{g}_l$ from the left channel are concatenated with the stream $\mathbf{k}_r$ from the right channel to form an ensemble stream $\mathbf{m}_l$ and the relationship between these streams are given as follows:

$$\mathbf{m}_l = [\mathbf{g}_l, \mathbf{k}_r] \qquad (8)$$

where $[\cdot, \cdot]$ means the concentration operation. The dimension space of the concatenated stream $\mathbf{m}_l$ is $m \times n \times (k + \omega)$ after the concentration operation conducted. It is clear that the information streams from the two channels are fused after the cross connection block conducted and the information stream

from the original main channel occupies a large proportion while the information stream from the other channel also transmits into the main channel via the cross connection block.

B. PROGNOSTIC MODEL CONSTRUCTION

The architecture of the CF-ConvLSTM is shown in Fig. 2 and the detailed parameters of the proposed model is listed in Tables 1-3. In the proposed model, two input channels are constructed to make a full utilization of the representative features from multiple monitoring data, a series of ConvLSTM cells are stacked to reflect the spatial-temporal correlation of the representative feature and the batch normalization (BN) [24] and leaky linear unit [23] are affiliated in each ConvLSTM cell to reduce the probability of vanishing gradient in the proposed network structure. Simultaneously, the proposed
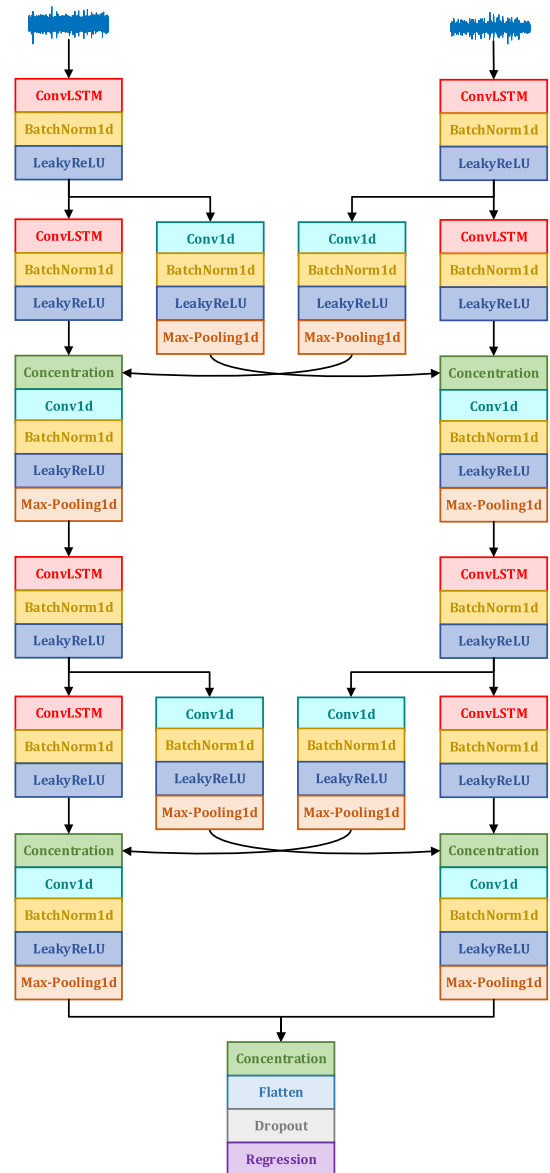

FIGURE 2. The schematic of the CF-ConvLSTM.

No.	Network layer	Input channel	Hidden channels	Kernel size	Step size
1	ConvLSTM1	1	[2,4]	5	2
2	ConvLSTM2	4	[6,8]	5	2
3	ConvLSTM3	8	[10,12]	5	2
4	ConvLSTM4	12	[14,16]	5	2

TABLE 2. The parameter setting of the convolution layers in the proposed network structure.

No.	Network layer	Input channel	Output channels	Kernel size	Stride size
1	Conv1	4	8	5	2
2	Conv2	12	16	5	2
3	Conv3	8	8	5	2
4	Conv4	16	16	5	2

TABLE 3. The parameter setting of the max-pooling layers in the proposed network structure.

No.	Network layer	Kernel size	Stride size	Padding size
1	Max-pooling1	3	2	1
2	Max-pooling2	3	2	1
3	Max-pooling3	3	2	1
4	Max-pooling4	3	2	1

cross connection block is imposed on the network structure twice to make a fusion of the information streams from two adjacent channels. In the end of the network structure, a concentration operation is adopted to integrate the information streams from two channels into an single form and then the representative features is flatted into one-dimension vector to import into a full connection layer to achieve the RUL predication. Meanwhile, a dropout operation [25] with a dropout rate 0.5 is used in the end of the network to improve the generalization ability of the proposed model.

Moreover, it should be noticed that due to the bearing RUL prediction can be formulated as a regression problem in essence, the mean squared error (MSE) between the prediction result $P = [p_1, \ldots, p_b, \ldots, p_B]$ and the actual reliability rate $R = [r_1, \ldots, r_b, \ldots, r_B]$ (where $b \in [1, 2, \ldots, B]$ means the number of recorded samples in the monitoring dataset) is introduced as the optimization objective in this study.

C. MULTI-AVERAGING APPROACH TO SMOOTH THE PREDICTION RESULT

Due to the data driven approach adopted in this study, there exists a large fluctuation on the prediction result, which make it inconvenient for directly observation. Thus, it is necessary to smooth the prediction result to make it less affect by the large fluctuation. In this study, a multi-averaging approach is proposed to smooth the prediction result and its main principle is to conduct the averaging operation on the prediction result multiple times. The averaging operation is implemented as follows:

$$\hat{p}_b = \begin{cases} mean\,(p_1, \ldots, p_b)\,, & \text{if } b \leqslant sl \\ mean\,(p_{b-sl}, \ldots, p_b)\,, & \text{if } b > sl \end{cases} \quad (9)$$

where sl denotes the section length for the first separated averaging operation in Eq. (9) and the numerical value 10 is

selected in this study after error and trials. It is clear that the prediction result is smoothed after conducting the averaging operation once though there still exists a fluctuation on the prediction result. The observation motivates us to conduct the averaging operation multiple time to largely eliminate the fluctuation in the prediction result and the full algorithm of the multi-averaging approach is given as follows:

Algorithm 1 The Procedure of the Smoothing Method

1. Input: Original prediction result $P = [p_1, \ldots, p_b, \ldots, p_B]$

    Iteration number for conducting multi averaging operation: Iter
2. **For** $e = 1$ **to** Iter **do**
3.   **For** $b = 1$ **to** B **do**
4.
$$\hat{p}_b = \begin{cases} mean\,(p_1, \ldots, p_b)\,, & \text{if } b \leqslant sl \\ mean\,(p_{b-sl}, \ldots, p_b)\,, & \text{if } b > sl \end{cases}$$

5.   **End**
6. $[p_1, \ldots, p_b, \ldots, p_B] = [\hat{p}_1, \ldots, \hat{p}_b, \ldots, \hat{p}_B]$
7. **End**

D. GENERAL DESCRIPTION OF THE PROPOSED METHOD
The flowchart of the proposed bearing RUL prediction method is presented in Fig. 3 and the main procedures of the proposed method is shown as follows:

1) Initialization: The monitoring data for the accelerometers mounted on two different directions are collected and the amplitude of the monitoring data is normalized into the range $[-1], [1]$ as follows:

$$\bar{y}_i = 2 * \frac{y_i - y\_min}{y\_max - y\_min} - 1 \quad (10)$$

where $Y = [y_1, y_2, \ldots, y_n]$ denotes the monitoring data, y_max and y_min are the maximum and minimum values of the recorded series, and correspondingly the normalized monitoring data is represented as $\bar{Y} = [\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n]$.

Meanwhile, a reliability rate $R = [R_1, R_2, \ldots, R_B]$ is assigned to each recorded samples as follows:

$$R_b = 1 - (1/B) \cdot b \quad (11)$$

where $b \in [1, 2, \ldots, B]$ means the number of recorded samples in the monitoring dataset. It is clear that the reliability rate linearly decreases from 1 to 0 with respect to the increment of the number of the recorded samples.

2) Model training: The monitoring data corresponding to several bearing degradation processes in together with its reliability rates are served as the training dataset to update the weight and bias of the network, and the network training process is not terminated until the maximum epoch is reached.
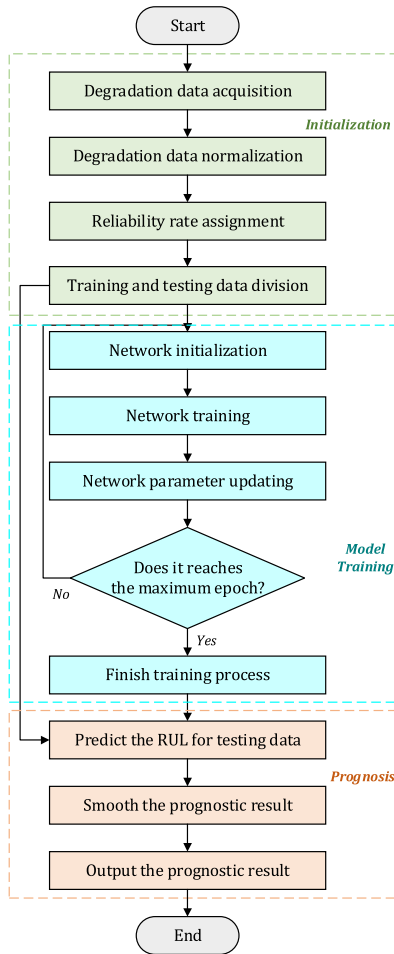
FIGURE 3. The flowchart of the proposed bearing RUL prediction method.

FIGURE 4. The schematic of the PRONOSTIA platform.

3) Performance evaluation: As the training process is finished, the monitoring data corresponding to the other bearing degradation process is used to test the performance of the trained prognostic model. Then the proposed smoothing operation is conducted on the prediction result to largely eliminate the fluctuation of the prediction result.

Furthermore, it is noted that the back-propagation is taken to update the weights of the proposed model and the stochastic gradient descent with momentum is used to minimize the loss function [26]. The learning rata is set as 0.00001 and the momentum term is set as 0.4. The training process is set as 100 and the batch size of the training samples is set as 7. The proposed algorithm is implemented using the PyTorch platform and NVIDIA Quadro P4000 GPU is adopted to take advantage of the GPU based accelerate computing architecture.

V. EXPERIMENTAL VERIFICATION
A. EXPERIMENTAL STEUP
The bearing degradation dataset comes from PRONOSTIA in the IEEE PHM 2012 Data Challenge is adopted in this
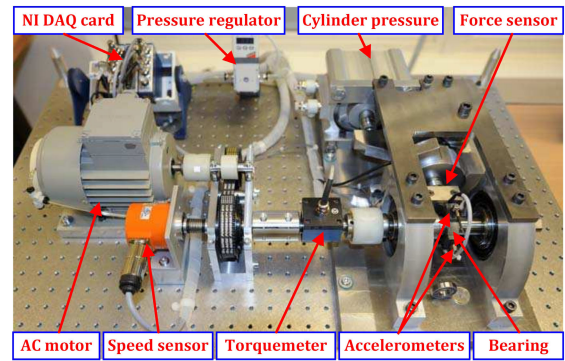
study [27]. The experiment platform is shown in Fig. 4. The PRONOSTIA platform is mainly composed of three parts, i.e., the rotating part, the degradation part and the monitoring data acquisition part. To accelerate the degradation process of the bearing, a constant radial load force is applied on the degradation part. Two accelerometers are mounted on the monitoring data acquisition part and these accelerometers are positioned at 90° to each other. The time length and the sampling frequency of each recorded monitoring data is 0.1s and 25600Hz. Thus, each recorded sample consists of 2560 data points. The sampling interval between two adjacent samples is 10s. Three working conditions are simulated in the degradation dataset with considering the changing of radial loads and operating speeds. In this study, the degradation dataset in the first working condition (the radial load and the operating speed in the first working condition is 4000N and 1800 rpm respectively, and the degradation data of sis tested bearing is recorded in this working condition) is used to verify the effectiveness of the proposed bearing RUL prediction method. Due to seven bearing degradation processes are simulated in the first dataset, each one is selected as the testing dataset and the other six ones are used to train the proposed model.

B. COMPARISON APPROACHES
In the proposed method, a CF-ConvLSTM model is proposed to achieve the bearing RUL prediction. To demonstrate the effectiveness of the main components of the proposed model, four comparison approaches are given and the details of these comparison approaches are given as follows: 1) Left- ConvLSTM: the cross connection block is removed and only the channel in the left hand is retained in this approach. 2) Right-ConvLSTM: the cross connection block is removed and only the channel in the right hand is retained in this approach. 3) CF-LSTM: the ConvLSTM cell is substituted as the LSTM cell in this approach. 4) CF-Conv: the ConvLSTM cell is substituted as the convolutional layer in this approach. It should be noticed that the first two approaches are constructed to confirm the effectiveness of the information fusion approach in the proposed model and the last two approaches are formulated to illustrate the usefulness of the ConvLSTM cell in the proposed model. In the first comparison approach,

the monitoring data acquired by the accelerator mounted on the horizontal direction of the bearing house is used. Meanwhile, the monitoring data acquired by the accelerator mounted on the vertical direction of the bearing house is selected in the second comparison approach. The monitoring data acquired by two accelerometers are adopted in the last two cases. The other detailed information (i.e., the setting of the training epochs, the learning rate and momentum term) of the comparison approaches is the same with that of the proposed model.

C. RESULTS AND DISCUEESIONS

To verify the effectiveness of the proposed multi-averaging approach in this study, the prediction result of bearing 3 (B3) obtained by the proposed method is given in Fig. 5. It can be found that there exists a large fluctuation in the original prediction result, which may lead to barrier for directly observation. To address this, the proposed smoothing method is applied on the original prediction result. It can be found that the fluctuation is largely eliminated when the averaging operation is conducted twice. Consecutively, when the iteration is conducted ten times, the smoothed result is flatter than the original prediction result, which is able to demonstrate the superiority of the proposed method.
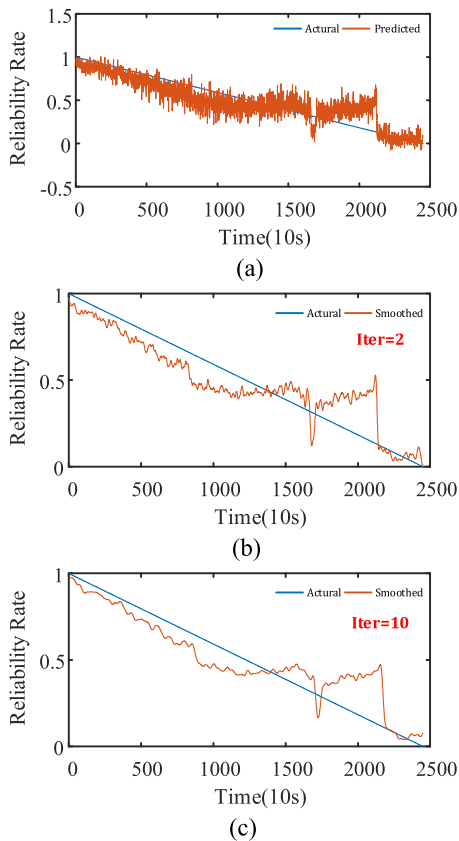
Moreover, to illustrate the effectiveness of the proposed method in achieving the prediction task, the prediction results provided by the proposed method and the other comparison methods listed in section 5.2 are given in Figs. 6-12. It should be noticed that to make a direct observation of the prediction results, all of them are smoothed via the proposed multi-averaging method and the iteration is set as 10. According to the prediction results, it is easier to see that all prediction results are clustered around the actual reliability rate, however, the distance between the prediction results
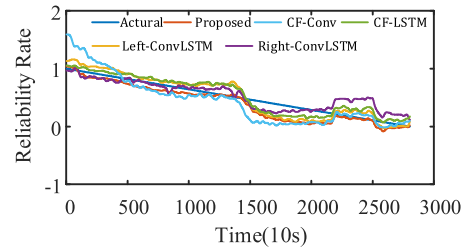

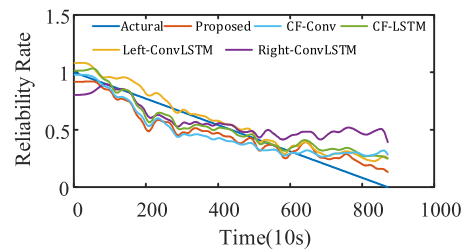**FIGURE 6.** The prediction results of bearing 1 (B1) obtained by different comparison approaches.

FIGURE 7. The prediction results of bearing 2 (B2) obtained by different comparison approaches.
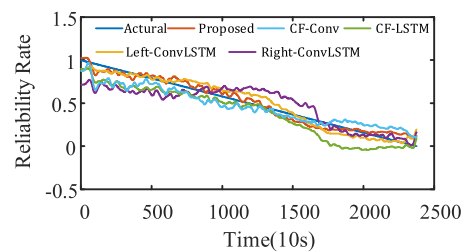

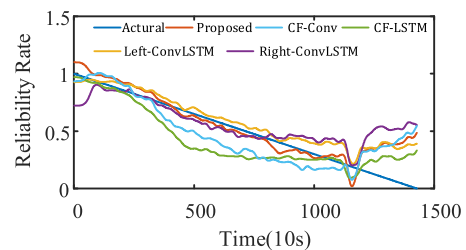**FIGURE 8.** The prediction results of bearing 3 (B3) obtained by different comparison approaches.

FIGURE 5. The prediction result of B3 obtained by the proposed method (a) the original prediction result, (b) the smoothed result when Iteration = 2 and (c) the smoothed result when Iteration = 10.

FIGURE 9. The prediction results of bearing 4 (B4) obtained by different comparison approaches.

TABLE 4. The prediction performance of different comparison approaches via two indices.

| No.	Indices	Proposed	Left-ConvLSTM	Right- ConvLSTM	CF-LSTM	CF-Conv
B1	RMSE(mean±std)	**0.1538±0.0005**	0.1602±0.0007	0.1717±0.0012	0.1545±0.0008	0.2254±0.0010
	Score(mean±std)	**26.9392±0.1019**	33.3354±0.1855	34.4356±0.2677	30.0911±0.1668	42.7017±0.2225
B2	RMSE(mean±std)	**0.1423±0.0009**	0.1925±0.0023	0.2380±0.0014	0.1886±0.0014	0.2324±0.0016
	Score(mean±std)	**8.9628±0.0639**	11.2069±0.1487	15.3143±0.1056	11.4580±0.1142	12.3404±0.1345
B3	RMSE(mean±std)	**0.1005±0.0007**	0.1713±0.0013	0.1926±0.0020	0.1137±0.0011	0.1678±0.0013
	Score(mean±std)	**17.1457±0.1357**	27.2897±0.2209	32.3522±0.3822	18.9007±0.2026	26.5460±0.2513
B4	RMSE(mean±std)	**0.1515±0.0007**	0.1777±0.0013	0.2218±0.0011	0.2179±0.0010	0.2428±0.0013
	Score(mean±std)	**14.8232±0.0753**	16.1228±0.1523	21.5502±0.1421	21.2193±0.1135	24.2958±0.1497
B5	RMSE(mean±std)	**0.1646±0.0005**	0.1704±0.0012	0.2134±0.0012	0.3119±0.0008	0.2215±0.0008
	Score(mean±std)	**29.0235±0.1163**	29.0800±0.2469	41.2101±0.2460	62.7202±0.1613	40.5417±0.1618
B6	RMSE(mean±std)	**0.1367±0.0007**	0.1580±0.0010	0.1724±0.0010	0.1725±0.0009	0.1889±0.0009
	Score(mean±std)	**22.9842±0.1516**	26.4438±0.2134	31.2287±0.1875	31.3998±0.1701	34.3681±0.1998
B7	RMSE(mean±std)	**0.1337±0.0007**	0.1672±0.0011	0.1727±0.0015	0.1794±0.0012	0.1888±0.0012
	Score(mean±std)	**22.4740±0.1368**	24.6638±0.1671	24.4126±0.2103	26.9260±0.2315	28.9458±0.2201
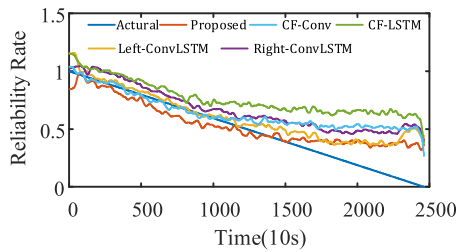
FIGURE 10. The prediction results of bearing 5 (B5) obtained by different comparison approaches.
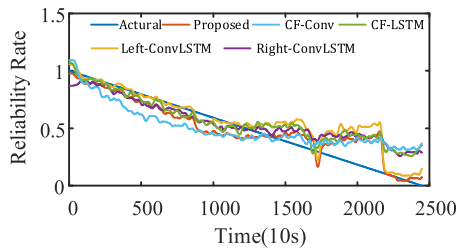

FIGURE 11. The prediction results of bearing 6 (B6) obtained by different comparison approaches.
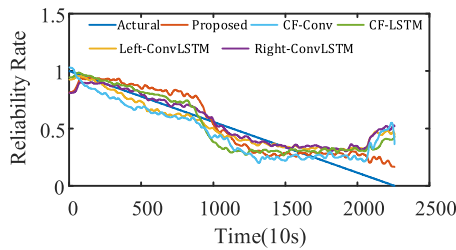

FIGURE 12. The prediction results of bearing 7 (B7) obtained by different comparison approaches.

obtained by the proposed method and the actual reliability rate is smallest among them. To make a quantitative analysis of the predication results obtained by different methods, two effectiveness indices widely adopted in the related references are introduced in this case, among them one is the root mean square error (RMSE), and its formula is given as follows:

$$RMSE = \sqrt{\frac{1}{B}\sum_{b=1}^{B}(p_b - r_b)} \tag{12}$$

The other is the scoring function originally formulated in the 2008 Prognostics and Health Management Data Challenge [28], which is defined as:

$$score = \begin{cases} \sum_{b=1}^{B}\left(e^{-\frac{d_1}{13}} - 1\right) & for\ d_b < 0 \\ \sum_{b=1}^{B}\left(e^{\frac{d_0}{10}} - 1\right) & for\ d_b \geqslant 0 \end{cases} \tag{13}$$

where $d_b = p_b - r_b$ denotes the error between the prediction result and the actual result. The scoring function is asymmetric so that the late predictions are more heavily penalized than early predictions.

The predication results obtained by different comparison methods are conducted 100 times, and a statistic of these two indices for the prediction results is given in Table 4. It can be seen that both the RMSE and score provided by the proposed method varies within a smallest range for all bearing datasets in comparison with those provided by the comparison methods listed in section 5.2, which indicates that the proposed method owns a great ability to predict the bearing RUL with high accuracy and stability.

VI. CONCLUSION

In this study, a cascade fusion architecture is proposed based on the foundation of the ConvLSTM cell, in which a cross connection block is formulated to make a fusion of the information from the adjacent channels twice and a concentration operation at the end of the two information streams is able to integrate them into one ensemble form. The proposed cascade fusion architecture can effectively make full use of the monitoring information from multiple sensory data, which owns a

great ability to predict the bearing RUL with a high accuracy and stability.

The validation on the experimental bearing degradation dataset provided by PRONOSTIA demonstrates that it is an effective approach to achieve the bearing RUL prediction task. The application of the multi-averaging approach on the initial prediction result is also helpful to make the prediction result smoothly. Moreover, the comparison with other data-driven RUL prediction methods can significantly highlight the superiority of the proposed method.

It is noted that the reliability rate is assumed to decrease linearly in this study, though there are also exist a number of assumptions on the variation of reliability rate. A reasonable assumption of the variation of reliability rate is also helpful to precisely achieve the RUL prediction task. Thus, we plan to further explore a more reasonable assumption of the variation of reliability rate in future work.

REFERENCES

[1] K. Yu, T. R. Lin, H. Ma, H. Li, and J. Zeng, "A combined polynomial chirplet transform and synchroextracting technique for analyzing nonstationary signals of rotating machinery," *IEEE Trans. Instrum. Meas.*, to be published, doi: 10.1109/tim.2019.2913058.

[2] G. Yu, "A concentrated time-frequency analysis tool for bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 371–381, Feb. 2020, doi: 10.1109/tim.2019.2901514.

[3] X. Yan, Y. Liu, M. Jia, and Y. Zhu, "A multi-stage hybrid fault diagnosis approach for rolling element bearing under various working conditions," *IEEE Access*, vol. 7, pp. 138426–138441, 2019.

[4] Z. Wang, L. Zheng, J. Wang, and W. Du, "Research on novel bearing fault diagnosis method based on improved krill herd algorithm and kernel extreme learning machine," *Complexity*, vol. 2019, pp. 1–19, Nov. 2019, doi: 10.1155/2019/4031795.

[5] I. El-Thalji and E. Jantunen, "Fault analysis of the wear fault development in rolling bearings," *Eng. Failure Anal.*, vol. 57, pp. 470–482, Nov. 2015.

[6] I. El-Thalji and E. Jantunen, "Dynamic modelling of wear evolution in rolling bearings," *Tribol. Int.*, vol. 84, pp. 90–99, Apr. 2015.

[7] K. Javed, R. Gouriveau, and N. Zerhouni, "A new multivariate approach for prognostics based on extreme learning machine and fuzzy clustering," *IEEE Trans. Cybern.*, vol. 45, no. 12, pp. 2626–2639, Dec. 2015.

[8] J. Ben Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, and F. Fnaiech, "Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network," *Mech. Syst. Signal Process.*, vols. 56–57, pp. 150–172, May 2015.

[9] J. Qu and M. J. Zuo, "An LSSVR-based algorithm for online system condition prognostics," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 6089–6102, Apr. 2012.

[10] S. Hong, Z. Zhou, E. Zio, and W. Wang, "An adaptive method for health trend prediction of rotating bearings," *Digit. Signal Process.*, vol. 35, pp. 117–123, Dec. 2014.

[11] J. Deutsch and D. He, "Using deep learning-based approach to predict remaining useful life of rotating components," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 1, pp. 11–20, Jan. 2018.

[12] K. Yu, T. R. Lin, and J. Tan, "A bearing fault and severity diagnostic technique using adaptive deep belief networks and Dempster–Shafer theory," *Struct. Health Monitor.*, vol. 19, no. 1, pp. 240–261, Jan. 2020.

[13] L. Ren, Y. Sun, J. Cui, and L. Zhang, "Bearing remaining useful life prediction based on deep autoencoder and deep neural networks," *J. Manuf. Syst.*, vol. 48, pp. 71–77, Jul. 2018.

[14] Q. Wang, B. Zhao, H. Ma, J. Chang, and G. Mao, "A method for rapidly evaluating reliability and predicting remaining useful life using two-dimensional convolutional neural network with signal conversion," *J. Mech. Sci. Technol.*, vol. 33, no. 6, pp. 2561–2571, Jun. 2019.

[15] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Rel. Eng. Syst. Saf.*, vol. 172, pp. 1–11, Apr. 2018.

[16] Y. Yoo and J.-G. Baek, "A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network," *Appl. Sci.*, vol. 8, no. 7, p. 1102, Jul. 2018.

[17] C.-G. Huang, H.-Z. Huang, and Y.-F. Li, "A bidirectional LSTM prognostics method under multiple operational conditions," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8792–8802, Nov. 2019.

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

[19] J. Chen, H. Jing, Y. Chang, and Q. Liu, "Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process," *Rel. Eng. Syst. Saf.*, vol. 185, pp. 372–382, May 2019.

[20] B. Wang, Y. Lei, N. Li, and T. Yan, "Deep separable convolutional network for remaining useful life prediction of machinery," *Mech. Syst. Signal Process.*, vol. 134, Dec. 2019, Art. no. 106330.

[21] J. Wu, K. Hu, and Y. Cheng, "Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network," *ISA Trans.*, to be published, doi: 10.1016/j.isatra.2019.07.004.

[22] X. Shi, Z. Chen, and H. Wang, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[23] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8609–8613.

[24] S. Loffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167

[25] N. Srivastava, G. Hinton, and A. Krizhevsky, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jan. 2014.

[26] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

[27] P. Nectoux, "Pronostia: An experimental platform for bearings accelerated degradation tests," in *Proc. IEEE Int. Conf. Prognostics Health Manag.*, 2012, pp. 1–8.

[28] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Proc. Int. Conf. Prognostics Health Manage.*, Oct. 2008, pp. 1–9.

QIONG WU received the B.A. degree in radio and television journalism from Liaoning Technical University, Fuxin, China, in 2015, and the M.Eng. degree in computer technology from Northeastern University, Shenyang, China, in 2019.

Her current research interests include deep learning, evolutionary computation, and their applications.

CHANGSHENG ZHANG received the Ph.D. degree in the computer science and technology from Jilin University, Changchun, China, in 2009.

He has authored two books and published more than 100 papers in international conferences and journals, which have been cited over one thousand times. His current research interests include evolutionary computation, distributed constraint programming, machine learning methods, and their applications. He is currently a Senior Member of the China Communication Society. In 2010, he was awarded excellent postdoctoral of Northeastern University, and his doctoral dissertation was awarded Excellent Doctoral Dissertation of Jilin Province, in 2011.

● ● ●