IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 28, 2021, accepted April 9, 2021, date of publication April 20, 2021, date of current version April 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3073945

A Framework for Predicting Remaining Useful Life Curve of Rolling Bearings Under Defect Progression Based on Neural Network and Bayesian Method

MASASHI KITAI[1,2], TAKUJI KOBAYASHI[2], HIROKI FUJIWARA[3], RYOJI TANI[3], MASAYUKI NUMAO[4], (Member, IEEE), AND KEN-ICHI FUKUI[4], (Member, IEEE)

[1]Graduate School of Information Science and Technology, Osaka University, Suita 565-0871, Japan
[2]NTN Next Generation Research Alliance Laboratories, Osaka University, Suita 565-0871, Japan
[3]NTN Corporation Advanced Technology Research and Development Center, Kuwana 511-0867, Japan
[4]Institute of Scientific and Industrial Research, Osaka University, Ibaraki 567-0047, Japan

Corresponding author: Masashi Kitai (kitai@arl.eng.osaka-u.ac.jp)

ABSTRACT In order to improve Remaining Useful Life (RUL) prediction accuracy for rolling bearings under defect progressing, the robustness for individual differences and the fluctuation of vibration features are challenging issues. In this research, we propose a novel RUL prediction framework based on a Convolutional Neural Network (CNN) and Hierarchical Bayesian Regression (HBR) for considering the degradation conditions and individual differences of RUL to improve the prediction accuracy. The characteristics of the proposed framework are: (1) In order to reduce the effect of the fluctuation of vibration features, the proposed framework uses an intermediate variable indicating the degradation condition instead of predicting RUL from vibration features. (2) The proposed framework considers not only present but also past degradation conditions in CNN. We conducted the experiment on rolling bearings under defect progression and evaluated the RUL prediction accuracy of the proposed framework. The proposed framework can generate a monotonous RUL prediction curve with a probability distribution and improve the RUL prediction accuracy under defect progression.

INDEX TERMS Convolutional neural network, feature fusion, hierarchical Bayesian regression, remaining useful life, rolling bearings.

I. INTRODUCTION

Rolling bearings are one of the essential mechanical elements in rotating machinery. In general, rolling bearings are often replaced when some kind of defect occurs on the raceway surface. However, in situations where it is not easy to replace the rolling bearing, where maintenance costs are high, or where somewhat defects on the rolling bearings do not interfere with the operation of the applications, the rolling bearing may be used even after defects have occurred. In such cases, it is necessary to predict the Remaining Useful Life (RUL) to the limit that can be used under defect progression. It is known that the lives of rolling bearings vary widely [1]. Furthermore, vibration acceleration features fluctuate greatly up and down under defect progression. Therefore, to create an appropriate maintenance plan, it is necessary to consider the individual differences in the RUL of each rolling bearing and fluctuation in vibration features and to guarantee the monotonicity of the RUL degradation.

There are two main approaches for predicting the RUL of rolling bearings: Time Based Maintenance (TBM) and Condition Based Maintenance (CBM). TBM is a concept of maintenance of rolling bearing as a whole group, and L_{10} life [2] is generally used. Given the variation of rolling bearing lives, the L_{10} life is calculated based on the Weibull distribution and is defined as the total rotation cycles or total operating time that 10 % of the rolling bearings are damaged when many rolling bearings (same type) are operated under the same conditions. TBM may require rolling bearings that are perfectly functional to be replaced, or serious defects may

The associate editor coordinating the review of this manuscript and approving it for publication was Long Wang.

occur before the periodic inspection, increasing the cost of maintenance.

On the other hand, CBM is a maintenance concept that considers the condition of each rolling bearing. Traditional CBM calculates the RUL for each rolling bearing by extracting the degradation index from the features of the vibration data and estimates the remaining operating time until the trend of the degradation index exceeds the threshold [3]. In recent years, several research reports have addressed CBM using vibration signals, contaminants in the lubricant, and the temperature and acoustic emissions of rolling bearings [4], [5]. Yet much of the previous work targets only early stage defect progression. Only a few studies have focused on estimating the relationship between the vibration signals and the defect progression [6], [7]. However, in both cases, these studies estimate the defect conditions and there is no mention of RUL.

Further, machine learning and deep learning methods have recently attracted attention as methods predicting RUL based on CBM. For example, RUL prediction models based on support vector machine or deep belief network were proposed to improve the prediction accuracy [8], [9]. On the other hand, there are also studies on RUL prediction methods based on the Bayesian estimation [10]. The regression models based on Bayesian estimation were proposed to predict the RUL curve and its reliability from the vibration features of rolling bearings [11], [12]. In addition, a regression model was proposed that takes into account RUL individual differences between each rolling bearing by hierarchizing the Bayesian regression (HBR) model [13]. However, these studies deal with the early stages of defect progression. Under the late stages, vibration fluctuation becomes larger. Therefore, we need an RUL prediction framework that suppresses the effects of vibration fluctuation under the late stages of defect progression and precisely predicts RUL.

To solve these problems, the authors previously proposed an RUL prediction framework combining Random Forest (RF) and HBR [14]. The characteristics of the previously proposed framework are as follows:

- By considering the relationship between defect progression and RUL instead of the relationship between operating time and vibration feature, the effect of the fluctuation of vibration feature was decreased.
- By using a Bayesian Regression (BR) model that inputs the circumferential defect length of the rolling bearing (defect size, DS) and outputs an RUL ***curve***, a monotonic decrease in the RUL prediction is guaranteed.
- To predict the RUL curve for a rolling bearing sample whose RUL and DS are unknown, the inputs (RUL and DS) for the BR model are estimated from vibration acceleration features using RF.
- By considering the individual differences of rolling bearings with HBR, the RUL prediction accuracy is improved, especially for the early stages of defect progression.

The RUL prediction accuracy was improved using the previously proposed framework. However, this framework has a problem in that the accuracies of predicting the DS and RUL by RF are not high enough, and this causes a decrease in the estimation accuracy of the BR model.

In this paper, we propose a novel RUL prediction method based on the Convolutional Neural Network (CNN), named Feature Fusion Network (FFN) to improve the prediction accuracy. The additional characteristics of FFN are as follows:

- By using CNN for the RUL prediction and using spectrograms of vibration acceleration as inputs, we can consider the spatial correlation of the spectrograms.
- By normalizing (min-max scaling) the RUL / DS and using them as supplemental information of FFN, degradation conditions can be considered.

With the proposed FFN, we can improve the RUL prediction accuracy in both the early and late stages of the defect progression. In this paper, we provide an RUL prediction framework that combines FFN with HBR. By using the prediction framework, we can guarantee the monotonicity of RUL degradation and consider the RUL individual differences of rolling bearings.

The rest of this paper is organized as follows. Section II introduces the background of this research. Section III describes the related works on RUL prediction for rolling bearings. The details of the proposed RUL prediction framework (Overview of the framework and detailed structure of FFN and HBR) are presented in Section IV. The experimental conditions, evaluation method, and results of the proposed RUL prediction framework are described in Section V. Finally, the conclusions are presented in Section VI.

II. PROBLEMS OF PREDICTING RUL UNDER DEFECT PROGRESSION

A. DEFECT PROGRESSION AND VIBRATION FEATURES

For rolling bearings used under unidirectional loading conditions, defect (flaking) often occurs on the raceway surface of the fixed ring (in this paper, the inner ring) of rolling bearings. If the operation is continued after the initial defect occurs on the raceway surface, the defect expands in axial and circumferential directions. This causes vibration acceleration to increase as the defect progresses. Additionally, the upward trend of vibration varies depending on the defect state. Fig. 1 shows the relationship between the operating time and Root-Mean-Square (RMS) value of vibration acceleration. It also shows the relationship between the operating time and defect length in the axial and circumferential directions, as well as the RUL from the measurement time to the operational limit. In this paper, the operational limit is assumed to be the time when the DS reaches 12 mm, due to sudden increment of vibration amplitude, it is difficult to continue the experiment for some bearing samples.

Fig. 2 shows an overview of the rolling bearing and the defect shape on the raceway of the inner ring. The right two

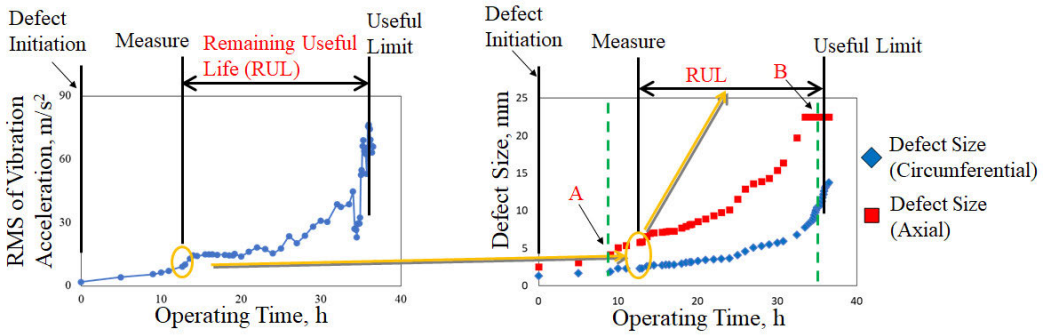

FIGURE 1. Relationship between operating time and vibration feature (left) and operating time and defect size (right) under damage progressing.

FIGURE 2. Overview of rolling bearing and defect shape on the raceway surface of the inner ring.

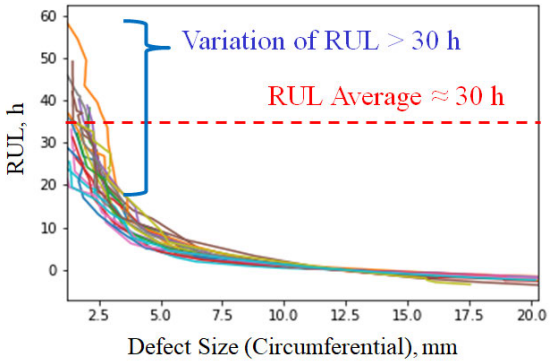

FIGURE 3. Variation of RUL among the samples.

images in Fig. 2 indicate the condition of defect progression at Points A and B in Fig. 1. The RMS value is generally used for rolling bearing diagnostics [3], but it varies greatly around the end of defect progression, and it is difficult to accurately determine the condition of rolling bearings. On the other hand, the change in axial and circumferential defect length is a suitable index for predicting the RUL because it shows a monotonous change with respect to operating time. However, it is difficult to measure the DS during operation because the equipment has to be stopped or disassembled.

B. VARIATION IN ROLLING BEARING RUL

The time until a defect occurs has variation among the rolling bearings in general. We need to consider the same problem for predicting RUL under defect progression. Fig. 3 shows the relationship between the DS and RUL until the DS reaches the specified length for 33 rolling bearing samples. Although the multiple rolling bearings shown in Fig. 3 are all measured under the same operating conditions, the variations in RUL are larger than the average RUL. Therefore, an RUL prediction method that takes variation into account is important for improving prediction accuracy.

III. RELATED WORK

Ren *et al.* [15] used min-max scaled RUL according to operating conditions as a kind of Health Indicator (HI) [16], and used Restricted Boltzmann Machine (RBM) and Gated Recurrent Unit (GRU) to predict the HI from vibration features of PRONOSTIA dataset [17]. Guo *et al.* [18] used min-max scaled operating time according to operating conditions as the HI and used Recurrent Neural Network (RNN) to predict the HI. Additionally, they estimated raw RUL from the HI for the vibration features of the PRONOSTIA dataset and Gearbox bearing dataset. Liu and Gryllias [19] used Bidirectional Long Short Term memory (Bi-LSTM) [20] and Domain Adversarial Neural Network (DANN) [21] to predict raw RUL from vibration features of the XJTU-SY [22] dataset.

In addition, vibration acceleration is most commonly used for rolling bearing diagnosis, and CNN has been attracting attention as a method for extracting useful features for diagnosis. Yoo and Baek [23] and Ren *et al.* [24] used quantified machine conditions or min-max scaled RUL according to operating conditions as HI. They predicted the HI using wavelet images or FFT-based images of vibration acceleration as inputs of 2D-CNN, respectively. Guo *et al.* [25] predicted the HI (min-max scaled operating time) by using the time-series data of vibration acceleration directly as inputs of 1D-CNN and improved the prediction accuracy by correcting the variation of the estimated result. Zhu *et al.* [26] predicted raw RUL by combining the features of different

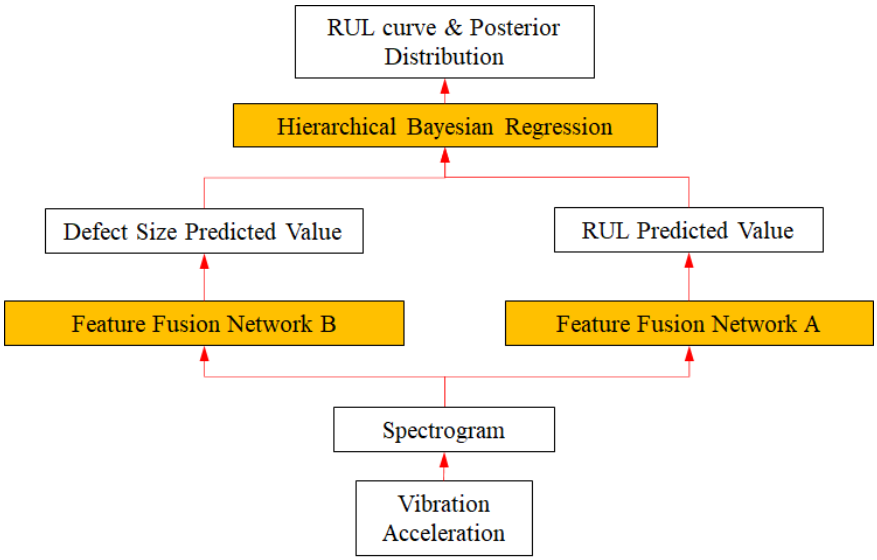

FIGURE 4. Flow of the proposed RUL prediction framework.

layers of 2D-CNN and using them as the input of the flatten layer. Hinchi and Tkiouat [27] and Jiang *et al.* [28] predicted raw RUL using CNN-LSTM and time-series of vibration acceleration or statistical features of vibration acceleration as inputs, respectively. Wang *et al.* [29] used a combined method of CNN and GRU to predict raw RUL from time-series of vibration acceleration and estimated RUL reliability using variational inference.

In the abovementioned papers, [15] and [25] used HI as an objective variable; however, they did not consider raw RUL. Also, [15], [18], [19], and [23]– [26] did not consider the past conditions of degradation, and RUL prediction error may increase if the fluctuation of vibration features becomes large. Then, [15], [27], [28], and [29] considered RUL degradation by using LSTM or GRU but did not consider its monotonicity. Furthermore, the abovementioned papers ([15], [18], [19], [23], and [25]- [29]) did not investigated the RUL of rolling bearings that are used even under the defect progression.

In summary, none of the existing studies take into account past degradation, individual differences, and monotonicity of RUL in a single framework.

IV. PROPOSED METHOD

A. OVERVIEW

We need to consider the fluctuation of vibration features and individual differences in RUL for each bearing sample to guarantee the monotonicity and robustness of the RUL prediction results. In addition, it is important to consider the degradation conditions to improve RUL prediction accuracy. In this paper, we propose the CNN-based RUL prediction framework named Feature Fusion Network (FFN) using both vibration acceleration spectrograms and past degradation conditions by using normalized RUL and DS to improve prediction accuracy. Furthermore, by combining FFN with HBR, we consider the fluctuation of vibration features and RUL individual differences to obtain the RUL regression ***curve*** and reliability for each bearing sample, respectively. Fig. 4 shows the overview of the proposed framework. FFN A and FFN B have the same network structure and use the same inputs (spectrogram of vibration acceleration). The only difference between the two FFNs is whether the objective variable is RUL or DS.

We assume that the training bearing samples have data of vibration acceleration, DS, and RUL at all measurement times. At the same time, we assume that the testing bearing sample has only vibration acceleration data until the measurement time. We predict the DS and RUL of the testing bearing samples using FFN. Also, we use these predicted DS and RUL as inputs of HBR to predict the RUL curve and its reliability (i.e., posterior distribution).

B. INPUT FEATURE

We use spectrograms of vibration acceleration as inputs to FFN A and FFN B. The spectrograms are converted from time-series waveform of vibration acceleration of using Short Time Fourier Transform (STFT). The spectrogram is obtained by cutting out a part of the signal (vibration acceleration) using the sliding window, then Fourier transform is applied to each cut out signal.

C. FEATURE FUSION NETWORK

1) STRUCTURE OF FEATURE FUSION NETWORK

Fig. 5 shows the overview of the FFN model. The FFN model consists of CNN Unit 1, CNN Unit 2, Degradation Index Vector (DIV) Unit, and Fusion Unit. CNN Units 1 and 2 are respectively formed by convolutional layers, pooling layers, batch normalization layers, and a flatten layer. The DIV Unit and Fusion Unit are respectively formed by fully connected layers. CNN Unit 1 extracts feature vectors ($\boldsymbol{f}_{conv}$) from spectrograms of vibration acceleration data on present time. CNN Unit 2 estimates normalized objective variables (normalized DS or normalized RUL) from

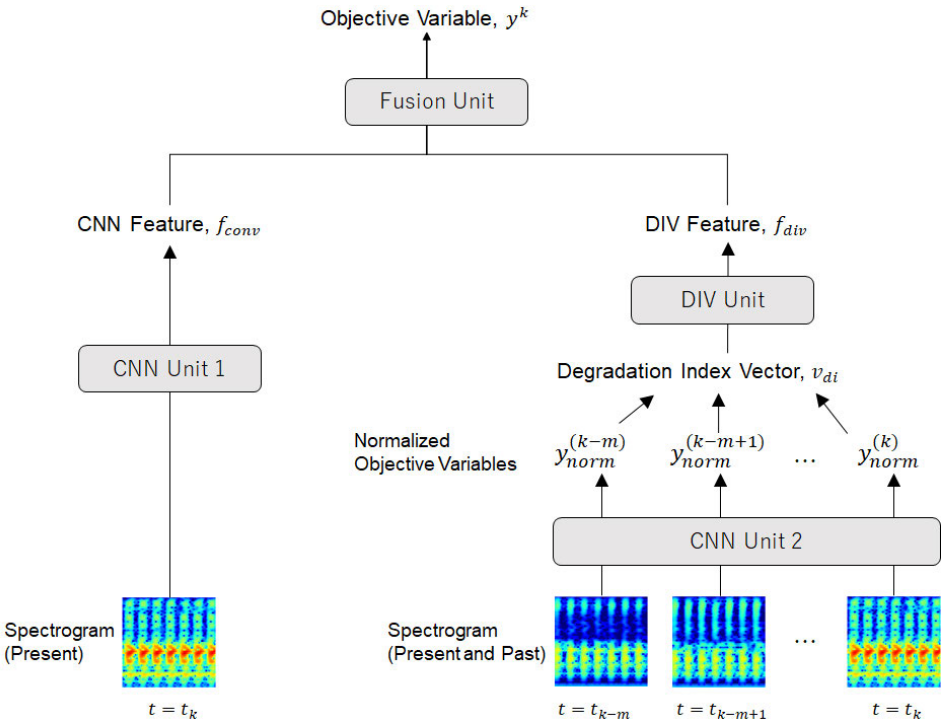

FIGURE 5. Feature fusion network.

spectrograms of vibration acceleration data on present and past times. DIV ($\boldsymbol{v}_{di}$, described in IV-C2) is obtained by vectorizing these normalized objective variables in measurement order. The DIV Unit estimates the feature vector ($\boldsymbol{f}_{div}$) from DIV ($\boldsymbol{v}_{di}$). The Fusion Unit combines $\boldsymbol{f}_{conv}$ and $\boldsymbol{f}_{div}$ and uses this as input to estimate the objective variables (RUL or DS). We aim to consider the degradation conditions of bearing and improve prediction accuracy using $\boldsymbol{v}_{di}$.

2) DEGRADATION INDEX VECTOR

$\boldsymbol{v}_{di}$ is a vector indicating the present and past degradation conditions of RUL degradation. It is used to improve the prediction accuracy of FFN. $\boldsymbol{v}_{di}$ consists of normalized objective variables (normalized RUL or normalized DS) y_{norm} vectorized in the order of measurement. y_{norm} is calculated by min-max normalization given by Eq. (1), and $\boldsymbol{v}_{di}$ is given by Eq. (2).

$$y_{norm}^{(k)} = \frac{y^{(lim)} - y^{(k)}}{y^{(lim)} - y^{(1)}} \tag{1}$$

$$\boldsymbol{v}_{di}^{(k)} = [y_{norm}^{(k)}, y_{norm}^{(k-1)}, \cdots, y_{norm}^{(k-l)}, \cdots, y_{norm}^{(k-m)}] \tag{2}$$

Here, $k = [1, 2, \cdots, lim]$ indicates the index of the measurement data of each bearing sample, and m indicates the number of past degradation conditions to be considered in $\boldsymbol{v}_{di}$ ($0 \leq m \ll lim$). When $m = 0$, $\boldsymbol{v}_{di}$ indicates only the current degradation condition. On the other hand, by increasing the value of $m\,(\geq 1)$, we can also consider past degradation conditions in $\boldsymbol{v}_{di}$. $y^{(k)}$ is the objective variable at measurement timing k. $y^{(1)}$ and $y^{(lim)}$ are the initial state and final state of the objective variable for each bearing sample, respectively. In Eq. (2), when $k - l < 1$, $y_{norm}^{(1)}$ is used instead of the past state because the past state does not exist.

For training bearing samples, y_{norm} can be calculated from a given y, but for testing bearing samples, the y is unknown. In this paper, we predict y_{norm} for testing bearing samples using CNN Unit 2 as described in Fig. 5 in advance and use the values to find $\boldsymbol{v}_{di}$.

D. HIERARCHICAL BAYESIAN REGRESSION MODEL

Equations (3) to (10) show the RUL regression equations and the probability distributions given to each parameter of the regression equation in the HBR model for the proposed framework. Level 1 below shows the relationship between the DS and the RUL of the i^{th} bearing sample. Level 2 shows the probability distributions of the common parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, individual parameter $\boldsymbol{\delta}_i$ for the i^{th} bearing sample, and parameters $\boldsymbol{\sigma}_y$ and $\boldsymbol{\nu}_y$, which indicate the uncertainty of the RUL. $\boldsymbol{\sigma}_y$ and $\boldsymbol{\nu}_y$ are the scale and degrees of freedom of the Student's t-distribution, respectively. $\boldsymbol{\sigma}_\delta$ is a hyperparameter of $\boldsymbol{\delta}_i$. Also, Level 3 shows the hyperprior distribution of $\boldsymbol{\sigma}_\delta$. $\boldsymbol{\delta}_i$ and $\boldsymbol{\sigma}_\delta$ are lognormal distributions since they do not take negative values. $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are normal distributions. The Student's t-distribution parameters $\boldsymbol{\sigma}_y$, $\boldsymbol{\nu}_y$ are half-Cauchy distribution and exponential distribution, respectively. Equations (5) to (10) show the prior distributions of each parameter, and we infer a posterior distribution of the parameters by the Markov Chain Monte Carlo algorithm in the training or testing phase, respectively.

Fig. 6 shows the conceptual diagram of the HBR model for the proposed framework. The invariant characteristics for all

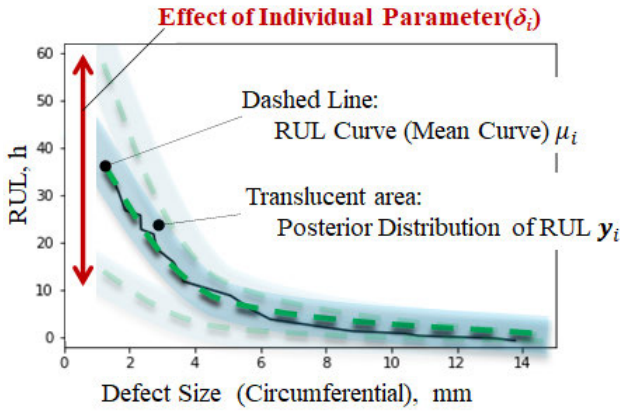

FIGURE 6. Concept of hierarchical Bayesian regression.

bearing samples are expressed by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and the individual characteristic for each bearing sample is expressed by $\boldsymbol{\delta}_i$. By giving $\boldsymbol{\sigma}_\delta$ as a hyperparameter of $\boldsymbol{\delta}_i$ and using HBR, the individual differences in the RUL can be expressed by one regression model.

1st Level

$$\boldsymbol{y}_i \sim StudentT(\boldsymbol{\mu}_i, \boldsymbol{\sigma}_y, \boldsymbol{\nu}_y) \tag{3}$$

$$\mu_i = \boldsymbol{\delta}_i(\boldsymbol{\alpha} + \frac{\boldsymbol{\beta}}{ds}) \tag{4}$$

2nd Level

$$\boldsymbol{\alpha} \sim Normal(0, 100) \tag{5}$$

$$\boldsymbol{\beta} \sim Normal(0, 100) \tag{6}$$

$$\boldsymbol{\sigma}_y \sim HalfCauchy(5) \tag{7}$$

$$\boldsymbol{\nu}_y \sim Exponential(0.03) \tag{8}$$

$$\boldsymbol{\delta}_i \sim Lognormal(0, \boldsymbol{\sigma}_\delta) \tag{9}$$

3rd Level

$$\boldsymbol{\sigma}_\delta \sim Lognormal(0, 100) \tag{10}$$

E. CALCULATION FLOW

The steps for predicting the RUL curve are as follows:

Step 1: Generating a normalized RUL and normalized DS for $\boldsymbol{\nu}_{di}$

The normalized RUL and normalized DS are predicted from spectrograms of vibration acceleration using CNN Unit 2. $\boldsymbol{\nu}_{di}$ of RUL and DS are constructed from these predicted normalized values, respectively. For testing bearing samples, normalized RUL and normalized DS are predicted by the CNN model, which is trained with all the training bearing samples.

Step 2: Predicting RUL and DS using FFN

The FFN model is trained using spectrograms and the $\boldsymbol{\nu}_{di}$ of RUL of all training samples, and the RUL of the testing bearing sample is predicted with the trained FFN. In the same way, the FFN model is trained using spectrograms and the $\boldsymbol{\nu}_{di}$ of the DS of all the training samples, and the DS of test bearing samples is predicted with the trained FFN.

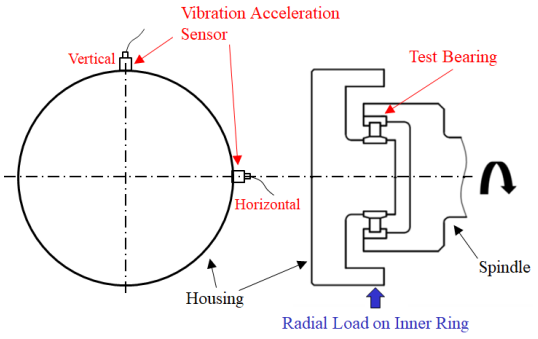

FIGURE 7. Test equipment.

Step 3: Inferring posterior distributions of common parameters and hyperparameters of HBR

The HBR model is trained using the true values of RUL and DS of the training bearing samples to infer $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}_\delta$.

Step 4: Predicting the RUL curve and posterior distributions using HBR

The predicted values of RUL and DS obtained in Step two and posterior distributions of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}_\delta$ are used to infer the posterior distributions of $\boldsymbol{\delta}_i$, $\boldsymbol{\sigma}_i$ and $\boldsymbol{\nu}_i$ for testing bearing samples using HBR. Finally, the RUL curve and posterior distribution of testing bearing samples are predicted using HBR and the parameters obtained above.

V. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL CONDITIONS

Fig. 7 shows a schematic diagram of the experimental equipment used to evaluate the RUL prediction accuracy, and Table 1 shows the main experimental conditions. Cylindrical roller bearings (Type: NU224) were used in the experiment, and the defect was considered to be the flaking on the raceway, which is the most common type of defect of rolling bearings.

Vibration acceleration (vertical and horizontal directions) and DS were measured every 20 minutes on average for 33 bearing samples. Data from the bearing samples were gathered from the normal condition until the defect progressed to the limit of operation, and the data after defect occurrence were used for evaluation. The time required for the defect to reach a specific size was taken as the reference point for RUL. The RUL was then determined by subtracting the operating time for each measurement from the reference. Examples of the measurement data and defect conditions are shown in Fig. 1 to Fig. 3 in Section 2.

One measurement of the vibration acceleration of one direction $\boldsymbol{X}^{(M)} = [x_1, x_2, \cdots, x_j, \cdots, x_N]$, $M \in \{\text{Vertical}, \text{Horizontal}\}$ was taken at a sampling frequency of 50 kHz and a sampling time of 20 seconds. Here, index j indicates the time-series order, and x_j indicates the instantaneous value of vibration acceleration amplitude at index j.

B. SHORT TIME FOURIER TRANSFORM

Table 2 shows the conditions of STFT. We preprocessed the vibration acceleration data with min-max scaling, STFT

TABLE 1. Experimental conditions.

Bearing	Cylindrical Roller Bearing (Type: NU224)
Rotational Speed	500 min^{-1}
Radial Load	90 kN
Measurement	Vibration Acceleration (Vertical and Horizontal)
# of Bearing Samples	33
# of Snapshots in Each Bearing Sample	51 - 129

processed under the conditions shown in the Table 2. Each of the spectrograms is resized as 128 × 128 pixels and used as the input to FFN.

C. EVALUATION MEASURE

The coefficient of determination (R^2) was used as an evaluation index of prediction accuracy. Equation (11) is the R^2 function.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{11}$$

y_i and $\hat{y}_i$ are the true value and predicted value, respectively. $\bar{y}$ is the average of the true values, and n indicates the number of data. R^2 was calculated for each bearing sample, and the average, median, and standard deviation of R^2 for all bearing samples were evaluated by leave-one (bearing)-out cross-validation. In other words, we selected one of the 33 bearing samples as the test sample, and used the remaining 32 bearing samples as the training samples. Then we evaluated R^2 scores for all of 33 test samples. The higher value and the less standard deviation of R^2, the better prediction result is. The variation of the R^2 score for each bearing is large due to individual differences of the degradation conditions and the difference of snapshot numbers for each bearing. Therefore, we focus on the prediction accuracy based on the median value. We also discuss the average and standard deviation in Section V-E2.

D. PROGRESSION STAGE FOR EVALUATION

The increasing ratio of defect progression speed and vibration amplitude change according to the state of defect progression. When the axial defect progression saturates, the circumferential defect progression speed increases (refer to Fig. 1). Therefore, we divided the progression stage into entire stage (S_{All}), early stage (S_{Early}), and late stage (S_{Late}) for the evaluation. Fig. 8 shows a schematic diagram of the evaluation stage. We visually checked the defect progression during the experiments, and we set S_{Early} and S_{Late} according to the axial defect progression.

E. PRELIMINARY EXPERIMENTS

1) PREDICTING NORMALIZED RUL AND NORMALIZED DS

This section confirms the effect of normalization of the objective variables to the prediction accuracy. Table 3 shows the parameters of the proposed FFN used for predicting RUL / DS. In this section, we use CNN Unit 2 in Table 3 to predict the raw or normalized RUL / DS. In the following

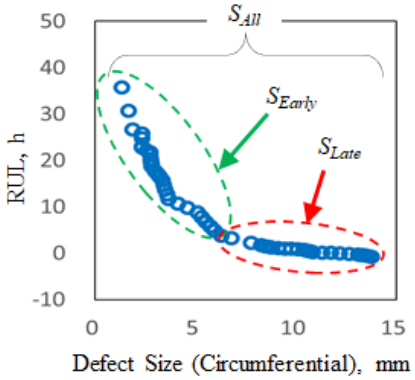

FIGURE 8. Stages for evaluating accuracy.

TABLE 2. Conditions of short time Fourier transform.

Data Length	4 s
Window Length	0.02 s
Overlap Ratio	0.05
Frequency Range	0 Hz - 8000 Hz
Window Function	Hanning

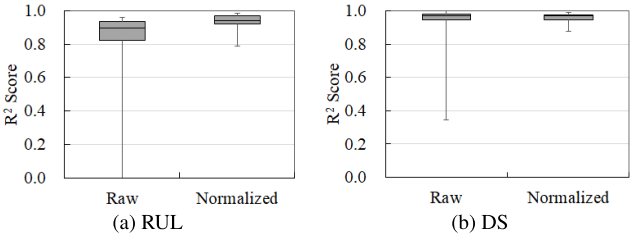

FIGURE 9. Effect of normalization on the prediction accuracy of RUL and DS.

sections, we use the validation dataset to determine the hyperparameters of the regression models.

Fig. 9 shows a comparison between the boxplots of the R^2 scores for the raw or normalized RUL / DS on the entire stage of defect progression. Due to normalizing the RUL, the median R^2 score can be improved from 0.82 to 0.93. More importantly, the variation in the R^2 scores becomes smaller, especially for RUL. By normalizing RUL / DS, R^2 scores are increased in specific bearing samples which have extremely low R^2 scores when predicting raw RUL / DS. Owing to normalizing RUL/DS, the variation between samples can be reduced, there is no extreme loss of accuracy for either normalized RUL or normalized DS. We confirm that the min-max normalization is effective for improving the prediction accuracy, especially for RUL.

2) EFFECT OF USING DEGRADATION INDEX VECTOR

FFN uses DIV to consider degradation conditions in CNN. Table 4 shows the averages, medians, and standard deviations of the R^2 score for the raw RUL when we select different m in FFN (w/o HBR). Table 4 also shows the average, medians, and standard deviation of the R^2 score of CNN using CNN Unit 2 in Table 3. For FFN, $m = 0$ indicates that FFN uses only the present degradation condition in v_{di}, and $m \geq 1$ indicates that FFN uses both present and past degradation conditions in v_{di}.

TABLE 3. Parameters of FFN.

Unit Name		Configuration
CNN Unit 1	Input	spectrogram of vibration acceleration (128 × 128 × 1)
	Layers	3 convolutions (filters: 16 - 8 - 4, kernels: 3 × 3, activations: ReLU) + pooling (kernel: 2 × 2) + batch normalization + dropout (dropout ratio: 0.5) + fully connected (units:1,024 - 100, activations: Linear)
	Output	f_{conv} (size,100)
CNN Unit 2	Input	spectrogram of vibration acceleration (128 × 128 × 1)
	Layers	3 convolutions (filters: 16 - 8 - 4, kernels: 3 × 3, activations: ReLU) + pooling (kernel: 2 × 2) + batch normalization + dropout (dropout ratio: 0.5) + fully connected (units:1,024 - 1, activations: Linear)
	Output	y_{norm} (size: 1)
DIV Unit	Input	v_{di} (constructed from y_{norm})
	Layers	fully connected (units:24 - 48, activations: Linear)+ dropout (dropout ratio: 0.5)
	Output	f_{div} (Size: 48)
Fusion Unit	Input	f_{conv} + f_{div}
	Layers	fully connected (unit:1, activation: Linear)
	Output	y (size: 1)

TABLE 4. Effect of the numbers of past degradation conditions (m) for FFN.

		Average			Median			Standard Deviation		
		S_{Early}	S_{Late}	S_{All}	S_{Early}	S_{Late}	S_{All}	S_{Early}	S_{Late}	S_{All}
CNN		0.554	-0.345	0.827	0.762	-0.106	0.898	0.707	0.931	0.232
FFN	m=0	0.597	0.353	0.846	0.761	0.534	0.921	0.742	0.591	0.241
	m=1	**0.651**	0.369	**0.863**	**0.821**	0.496	0.925	**0.504**	0.530	**0.170**
	m=3	0.625	**0.467**	0.854	0.801	**0.571**	0.920	0.623	0.398	0.207
	m=7	0.633	0.402	0.858	0.805	0.472	0.922	0.592	**0.396**	0.197
	m=15	0.587	0.395	0.844	0.805	0.497	**0.932**	0.837	0.441	0.269

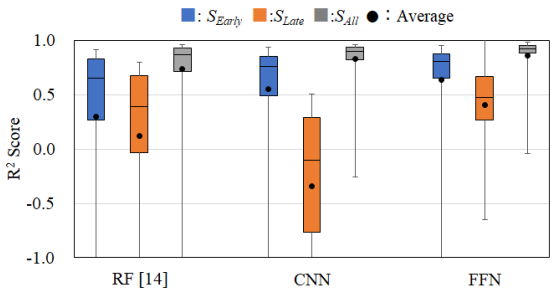

FIGURE 10. Prediction accuracy of RUL for each prediction method (w/o HBR).

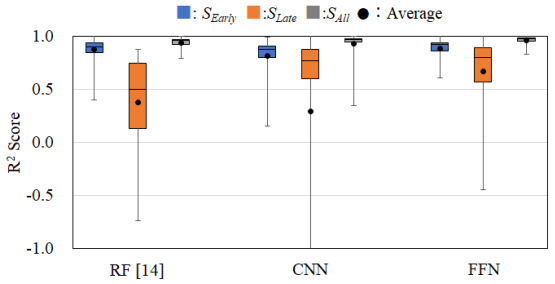

FIGURE 11. Prediction accuracy of DS for each prediction method (w/o HBR).

The median R^2 score of FFN in S_{Late} is improved about 0.6 compared to CNN just by the addition of considering the present degradation condition ($m = 0$). The median R^2 score in S_{Early} is improved when considering one past degradation condition ($m = 1$). On the other hand, the median R^2 score in S_{Late} has the highest value when three past degradation conditions ($m = 3$) are considered. From Fig. 8, in S_{Early}, the RUL value decreases rapidly as DS increases. In comparison, in S_{Late}, the RUL value slowly decreases compared to S_{Early}. Therefore, longer past data is needed to take into account the degradation trend in S_{Late}. The prediction accuracy is improved by referring to only one past condition in S_{Early}, and multiple (in this case, 3) past conditions in S_{Late}. By considering min-max scaled objective variables of present and past degradation conditions as DIV and adding them to CNN, we can improve the prediction accuracy for both S_{Early} and S_{Late}.

3) COMPARISON BETWEEN FEATURE FUSION NETWORK AND OTHER REGRESSION METHODS

This section compares the prediction accuracy between FFN (excluding HBR in the proposed framework), RF, and CNN. Table 3 shows the parameters of FFN used for predicting RUL and DS. The predicted normalized RUL and normalized DS (outputs of CNN Unit 2) are vectorized by Eq. (2) with $m = 7$ and used as the input(v_{di}) of the DIV unit.

In comparison, RF uses feature vectors that consist of 252 features considering all the combinations of band-pass filtering, domains (time, frequency, quefrency), sensor directions, and statistical features (such as Root-Mean-Square Value, Max Value, Kurtosis, and so on) [14]. CNN uses spectrogram as inputs, and the hyperparameters of CNN Unit 2 are the same as in Table 3. Note that RF and CNN do not consider degradation conditions (DIV).

Fig. 10 shows the comparison results of the R^2 score between RF, CNN, and FFN for RUL. For RUL prediction in RF, the average R^2 scores in S_{Early} and S_{Late} are lower than the median R^2 scores since some bearing samples have fairly low prediction accuracy. CNN can improve the prediction accuracy for those samples with low R^2 scores. Hence, both the median and average R^2 score can be improved using CNN. However, the prediction accuracy in S_{Late} of CNN becomes significantly lower than that of RF. Compared to RF and CNN, FFN can improve median and average R^2 scores in both

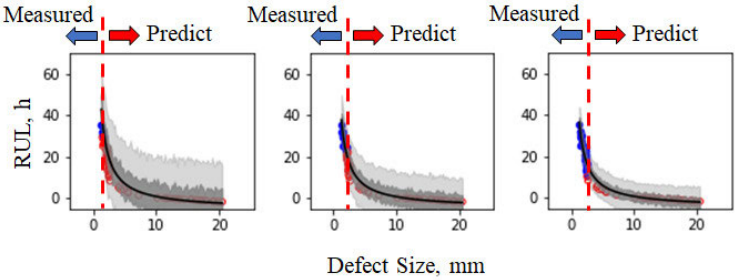

FIGURE 12. Relation between damage progression and RUL posterior distribution: Blue plots are measured values, and red plots are true values. The black line is the mean curve of prediction results, and the dark gray and light gray areas are 50% and 95% credible intervals, respectively.

of S_{Eearly} and S_{Late} by considering the degradation conditions as v_{di}. Especially, the median R^2 score of FFN in S_{Late} is about 0.25 higher than that of RF and 0.7 higher than that of CNN. Additionally, in every stage (S_{Early}, S_{Late}, S_{All}), the variation of the results from FFN becomes smaller than that from CNN and RF.

Similarly, for DS prediction (Fig. 11), we can improve the prediction accuracies using FFN compared to RF or CNN in every stage (S_{Eealy}, S_{Late},S_{All}). Especially in S_{Late}, FFN increases the R^2 score in specific bearing samples which have low R^2 scores with CNN, so that the average R^2 score of FFN considerably improved compared to CNN. The median R^2 score of FFN is about 0.3 higher than that of RF and 0.03 higher than that of CNN in S_{Late} for DS prediction.

Thus, we confirm that the RUL prediction accuracy for each stage of defect progression is improved by the proposed FFN.

F. EVALUATION OF PROPOSED FRAMEWORK

1) RELATIONSHIP BETWEEN DEFECT PROGRESS AND RUL CURVE

Fig. 12 shows the relationship between defect progression and the mean curve and posterior distribution of RUL predicted by the proposed framework. In Fig. 12, the blue plots indicate the measured value in the range used for training of HBR, and the red plots indicate the measured value in the range for prediction. The black line is the mean curve of the prediction result, and the dark gray and light gray areas are 50 % and 95 % credible intervals, respectively. As the measured data used for HBR training increases depending on the defect progression for the test sample, the mean curve approaches the true value and the credible intervals become narrow. By using HBR, the RUL can be expressed as the monotonous degradation curve with a probability distribution, and as the measurement proceeds, the prediction accuracy of the RUL can be increased along with the reduction of the interval.

2) PREDICTION ACCURACY OF THE PROPOSED FRAMEWORK

Fig. 13 shows the R^2 of RUL by the proposed framework (FFN+HBR). In Fig. 13, RF+HBR and CNN+HBR

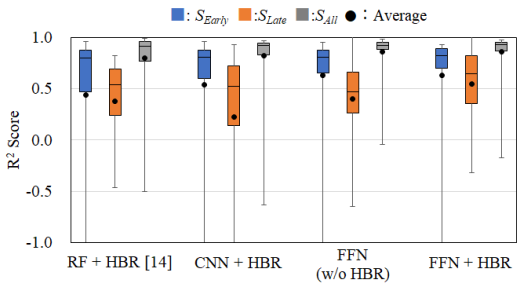

FIGURE 13. Prediction accuracy of the proposed framework.

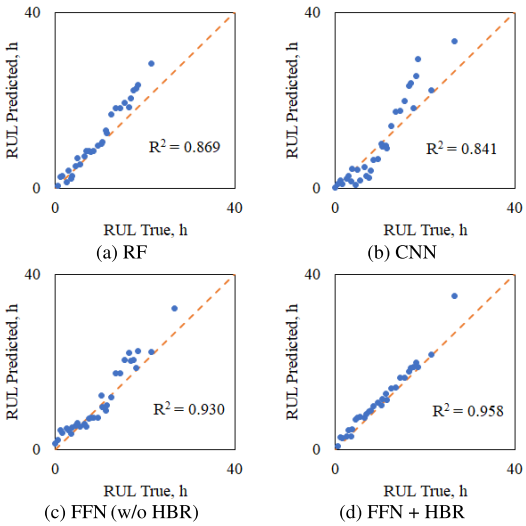

FIGURE 14. Correlation between the predicted value and the true value.

indicate the results using RF [14] or CNN (using only CNN Unit 1 in Table 3 and replacing the last fully connected layer with the output layer) in the proposed framework, respectively. FFN (w/o HBR) is the result using only FFN, and it is the same as in Section V-E3. The parameter settings of RF, CNN, and FFN are the same as in Section V-E3.

The median R^2 score of the proposed framework in S_{Early} is almost the same as for the other methods; however, the average R^2 score is about 0.2 higher than that of RF+HBR and 0.1 higher than that of CNN+HBR. As shown in Fig. 10, the snapshot RUL prediction accuracy of FFN is increased compared to RF and CNN. Thus, the RUL prediction accuracy in S_{Early} of the proposed framework (FFN+HBR) is improved compared to RF+HBR and CNN+HBR by

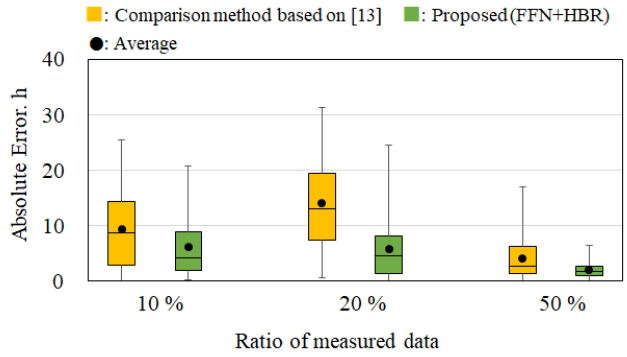

FIGURE 15. Comparison between the proposed framework and the method based on [13].

considering the past degradation conditions with DIV in FFN. Compared to FFN, the R^2 score of the proposed framework in S_{Early} is slightly improved (0.04 in median R^2 score, and 0.05 in average R^2 score) and the variation becomes smaller. Besides, the median R^2 score of FFN+HBR in S_{Late} is 0.12 higher than that of FFN. The average and min R^2 scores are also improved compared to FFN. The proposed framework can improve the prediction accuracy with consideration of individual differences and monotonicity of RUL.

Fig. 14 shows the results of the comparison of the predicted RUL values and the true RUL values for a specific bearing sample between RF, CNN, FFN, and the proposed framework (FFN+HBR). By considering the degradation condition with FFN, and considering the individual differences and monotonicity of RUL with HBR, we can suppress the variation in prediction result and improve the RUL prediction accuracy for each stage of defect progression and evaluate the probability distribution of RUL.

3) COMPARISON TO OTHER HBR-BASED METHOD

We compared the RUL prediction accuracy between the proposed framework (FFN+HBR) and the comparison method using HBR based on [13]. In the comparison method, RMS curve on the values of vibration acceleration, instead of RUL, is estimated by in which individual difference is modeled similar to the proposed method. Therefore, to calculate the RUL in the comparison method, we set the threshold by the RMS value at the operational limit (DS = 12 mm) for each bearing sample. Then, we estimate the RUL by the time that the regression curve exceeds the threshold. Additionally, it is difficult to calculate R^2 score because the calculation time explodes when we adopt the comparison HBR model to our experimental data. Therefore, we used absolute error, instead of R^2 score, at the timing when the ratio of measurement data becomes 10 %, 20 %, or 50 % of the operational limit for each bearing sample.

Fig. 15 shows the comparison between the boxplots of the absolute error for the comparison method and the proposed framework. In the proposed framework, the absolute errors are smaller than those of the comparison method 4.6 hours in average. Especially at 20 %, the median of absolute error was reduced about 8.5 hours. At this point, the average of the true RUL is about 26.6 hours, so the proposed framework improved the error rate about 32 %.

As shown in Fig. 1, the increment of vibration acceleration is non-monotonic, stagnates or fluctuates under defect progression. Prediction accuracy in the comparison method decreases according to the non-monotonicity of vibration acceleration increment since the comparison method uses the amplitudes of vibration acceleration to the output. The proposed framework can improve the prediction accuracy for each measurement timing.

VI. CONCLUSION

For predicting the RUL of rolling bearings under defect progression, we propose the Feature Fusion Network (FFN) that combines the Fully Connected Network using Degradation Index Vector (DIV) with Convolutional Neural Network (CNN). Additionally, we provide an RUL prediction framework that consists of FFN and Hierarchical Bayesian Regression (HBR) in order to guarantee the monotonicity and individual difference of rolling bearing RUL. The study findings are summarized below.

- The prediction accuracy of CNN (w/o DIV and HBR) was significantly improved by setting min-max scaled RUL/DS instead of raw RUL/DS as the objective variable, especially for bearing samples whose R^2 values were very low when raw RUL/DS was set as the objective variable.
- By considering not only present but also past min-max scaled RUL/DS as the degradation conditions and using them for FFN, we can improve the prediction accuracy for both the early and late stages of defect progression.
- By using the proposed framework combining FFN and HBR, we obtain higher R^2 scores than with FFN (w/o HBR) or other regression methods.

REFERENCES

[1] G. Lundberg and A. Palmgren, "Dynamic capacity of rolling beaings," *Acta Polytech.*, vol. 1, no. 3, pp. 3–50, 1947.
[2] T. A. Harris and M. N. Kotzalas, "Fatigue life: Basic theory and rating standards," in *Rolling Bearing Analysis*, 5th ed. New York, NY, USA: Taylor & Francis, 2006, ch. 11, pp. 195–252.
[3] L. Ma, J. S. Kang, and C. Y. Zhao, "Research on condition monitoring of bearing health using vibration data," *Appl. Mech. Mater.*, vols. 226–228, pp. 340–344, Nov. 2012.
[4] N. Jammu and P. Kankar, "A review on prognosis of rolling element bearings," *Int. J. Eng. Sci. Technol.*, vol. 3, no. 10, pp. 7497–7503, 2011.
[5] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, May 2018.
[6] Y. Li, C. Zhang, T. R. Kurfess, S. Danyluk, and S. Y. Liang, "Diagnostics and prognostics of a single surface defect on roller bearings," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 214, no. 9, pp. 1173–1185, Sep. 2000.
[7] H. Qiu, J. Lee, J. Lin, and G. Yu, "Robust performance degradation assessment methods for enhanced rolling element bearing prognostics," *Adv. Eng. Informat.*, vol. 17, nos. 3–4, pp. 127–140, Jul. 2003.
[8] Z. Liu, M. J. Zuo, and Y. Qin, "Remaining useful life prediction of rolling element bearings based on health state assessment," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 230, no. 2, pp. 314–330, Feb. 2016.
[9] 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.
[10] M. S. Hamada, A. G. Wilson, C. S. Reese, and H. F. Martz, *Bayesian Reliability*. New York, NY, USA: Springer, 2008, pp. 203–317.

[11] N. Z. Gebraeel, M. A. Lawley, R. Li, and J. K. Ryan, "Residual-life distributions from component degradation signals: A Bayesian approach," *IIE Trans.*, vol. 37, no. 6, pp. 543–557, Jun. 2005.
[12] R. Zhou, N. Z. Gebraeel, and N. Serban, "Degradation modeling and monitoring of truncated degradation signals," *IIE Trans.*, vol. 44, no. 9, pp. 792–803, 2012.
[13] M. Mishra, "Prognostics and health management of engineering systems for operation and maintenance optimisation," Ph.D. dissertation, Dept. Luleå, Univ. Tech., Luleå, Sweden, 2018, pp. 45–57. [Online]. Available: http://www.diva-portal.org/smash/get/diva2:1197741/FULLTEXT01.pdf and https://www.researchgate.net/publication/336022442_Prognostics_and_Health_Management_of_Engineering_Systems_for_Operation_and_Maintenance_Optimisation
[14] M. Kitai, Y. Akamatsu, R. Tani, H. Fujiwara, M. Numao, and K. Fukui, "Remaining useful life curve prediction of rolling bearings under defect progression based on hierarchical Bayesian regression," in *Proc. PAIS ECAI*, Santiago, Spain, Aug./Sep. 2020, pp. 2986–2992.
[15] L. Ren, X. Cheng, X. Wang, J. Cui, and L. Zhang, "Multi-scale dense gate recurrent unit networks for bearing remaining useful life prediction," *Future Gener. Comput. Syst.*, vol. 94, pp. 601–609, May 2019.
[16] D. Wang, K. Tsui, and Q. Miao, "Prognostics and health management: A review of vibration based bearing and gear health indicators," *IEEE Access*, vol. 6, pp. 665–676, 2018.
[17] P. Nectoux, R. Gouriveau, L. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Vanier, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in *Proc. Prog. Syst. Health Manag. Conf. (PHM)*, Saint Paul, MN, USA, Sep. 2012, pp. 1–8.
[18] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, May 2017.
[19] C. Liu and K. Gryllias, "Unsupervised domain adaptation based remaining useful life prediction of rolling element bearings," in *Proc. Prog. Syst. Health Manage. Conf.*, 2020, vol. 5, no. 1, pp. 1–10.
[20] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.
[21] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Machand, and V. Lempitsky, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 59, pp. 1–35, 2016.
[22] B. Wang, P. Vardon, and M. A. Hicks, "Rainfall-induced slope collapse with coupled material point method," *Eng. Geol.*, vol. 239, pp. 1–18, May 2018.
[23] 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, pp. 1–12, 2018.
[24] L. Ren, Y. Sun, H. Wang, and L. Xhang, "Prediction of bearings remaining useful life with deep covolutional neural network," *IEEE Access*, vol. 6, pp. 13041–13049, 2018.
[25] L. Guo, Y. Lei, N. Li, T. Yan, and N. Li, "Machinery health indicator construction based on convolutional neural networks considering trend burr," *Neurocomputing*, vol. 292, pp. 142–150, May 2018.
[26] 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. 2019.
[27] A. Z. Hinchi and M. Tkiouat, "Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network," *Procedia Comput. Sci.*, vol. 127, pp. 123–132, Mar. 2018.
[28] J. R. Jiang, J. E. Lee, and Y. M. Zeng, "Time series multiple channel convolutional neural network with attention-based long short-term memory for predicting bearing remaining useful life," *Sensors*, vol. 20, no. 166, pp. 1–19, 2020.
[29] B. Wang, Y. Lei, T. Yan, N. Li, and L. Guo, "Recurrent convolutional neural network: A new framework for remaining useful life prediction of machinery," *Neurocomputing*, vol. 379, pp. 117–129, Feb. 2020.

MASASHI KITAI received the master's degree in mechanical engineering from Hiroshima University, Japan, in 2009. He is currently pursuing the Ph.D. degree with the Graduate School of Information Science and Technology, Osaka University, Japan. Since 2009, he has been working with NTN Corporation, Japan. His research interests include artificial intelligence, machine learning, and their applications to machinery diagnoses.

TAKUJI KOBAYASHI received the B.E. and Ph.D. degrees in mechanical engineering from Kyoto University, Japan, and the M.S. degree in mechanical engineering from Northwestern University, USA. He is currently a specially appointed Professor with the NTN Next Generation Research Alliance Laboratories, Graduate School of Engineering, Osaka University, Japan. His research interests include fluid mechanics, molecular science, and machine learning with their applications to lubrication and bearing practice.

HIROKI FUJIWARA received the Ph.D. degree in engineering from Osaka University, Japan, in 2010. Since 1997, he has been working with NTN Corporation, Japan. His research interests include characteristics of rolling bearings, especially lubrication, contact problem, and vibration.

RYOJI TANI graduated from the Komono High School, Japan, in 2013. Since 2013, he has been working with NTN Corporation. His research interests include vibration and surface properties of rolling bearings.

MASAYUKI NUMAO (Member, IEEE) received the B.Eng. degree in electrical and electronics engineering, and the Ph.D. degree in computer science from the Tokyo Institute of Technology, Japan, in 1982 and 1987, respectively. He worked with the Department of Computer Science, Tokyo Institute of Technology, from 1987 to 2003, and a Visiting Scholar with CSLI, Stanford University, USA, from 1989 to 1990. He is currently a Professor with the Department of Architecture for Intelligence, Institute of Scientific and Industrial Research (ISIR), Osaka University, Japan. His research interests include artificial intelligence, machine learning, affective computing, and empathic computing. He is a member of the Information Processing Society of Japan, JSAI, the Japanese Cognitive Science Society, the Japan Society for Software Science and Technology, and the American Association for Artificial Intelligence.

KEN-ICHI FUKUI (Member, IEEE) received the master's degree from Nagoya University, Japan, in 2003, and the Ph.D. degree in information science from Osaka University, Japan, in 2010. From 2005 to 2010, he was a specially appointed Assistant Professor and an Assistant Professor with the Institute of Scientific and Industrial Research (ISIR), Osaka University, Japan, from 2010 to 2015. Since 2015, he has been an Associate Professor with the Institute of Scientific and Industrial Research (ISIR), Osaka University. His research interests include machine learning and data mining algorithms and its environmental contribution.