공학석사 학위논문

구름요소 베어링 진단을 위한 딥러닝 기반 스폴 크기 분포 추정 연구

A Study of Deep Learning-based Spall Size Distribution Estimation for Rolling Element Bearing Diagnosis

2023 년 2 월

서울대학교 대학원

기계공학부

정 화 용

구름요소 베어링 진단을 위한 딥러닝 기반 스폴 크기 분포 추정 연구

지도교수 윤 병 동

이 논문을 공학석사 학위논문으로 제출함
2022 년 10 월

서울대학교 대학원
기계공학부
정 화 용

정화용의 공학석사 학위논문을 인준함
2022 년 12 월

위 원 장 김 도 년 (인)

부위원장 윤 병 동 (인)

위 원 양 진 규 (인)

Abstract

A Study of Deep Learning-based Spall Size Distribution Estimation for Rolling Element Bearing Diagnosis

Hwayong Jeong

Department of Mechanical Engineering

The Graduate School

Seoul National University

When a rolling element bearing (REB) fails, the most common reason is the spall caused by rolling contact fatigue. In previous studies, when a ball passes through a spall, a step response with a low-frequency appears due to the effect of entering to the spall and an impulse response with a high-frequency appears when exiting the spall in the acceleration signal. Since the entry event signal is relatively weaker than the exit event signal and noise, research to date have attempted to estimate the location of the entry event using various signal processing technic such as noise reduction and strengthening the entry event features. However, in signal processing, manual parameter selection for finding the characteristics of entry event varies on bearing geometry and operating condition and since the parameter selection is empirical, the accuracy may differ accordingly. In addition, the spall size reflected in the signal also has uncertainty due to the geometry of the real spall and the uncertainty of rotation due to random slip. To overcome this difficulty, a deep learning-based approach was proposed in this study. The proposed architecture learned through analytic simulation signals which was generated by similar

geometry and operating conditions to test data, the spall size was estimated without manual parameter selection. By obtaining the mean and variance from the estimated values obtained from the models trained with several kernels and strides, the spall size distribution was obtained. The proposed method was validated through experimental data. Through the performance analysis results, the proposed method was effective.

Keywords : Spall size estimation
Rolling element bearing
Fault diagnosis
Denoising Autoencoder
Convolutional neural network
Ensemble learning

Student Number : 2018-27800

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Introduction

Rolling element bearing (REB) is a major element of a rotating machinery and its failure is the main reason for failure of the system[1][2][3]. By estimating the severity of bearing fault through condition monitoring, not only can major accidents be prevented in advance, but also appropriate maintenance strategies can be established to minimize maintenance costs.

The failure modes of REBs are classified into several categories, such as rolling contact fatigue, wear and plastic deformation and there are many reasons why they occur[4]. However, the most common reason of failure in well-installed, well-made bearings is spall due to rolling contact fatigue and the spall grows as the fault progresses. Therefore, spall size estimation can be a good means of severity estimation. By Epps[5] and Dowling[6], two main features originating from the rolling element passing through the spall were reported. The first originates from the entry event when the ball enter the spall, while the second results from the exit event when the ball contacts the edge of the spall. As the size of the spall increases, the separation between the entry and exit event increases and if the events can be extracted from the acceleration signal, the size of the fault can be estimated correctly. In actually, however, the events are weak and easy to be masked by noise.

Numerous studies have attempted to find the exact location of the entry-exit events. Sawalhi et al. used an autoregressive (AR) model as a pre-whitening for enhancing the energy of entry event relative to the exit event and a frequency band

selection with the complex Morlet wevelets[7]. Smith et al. suggested a low-pass filtering and steep roll-off to find the entry event point and envelope analysis is utilized to recognize the impact[8]. Cui et al. suggested the matching pursuit method to find the location of the entry and exit events by a step-impulse atom[9]. Chen et al. used variational mode decomposition (VMD) to separate the entry and exit event and used cross-correlation to find the spall size[10]. Zhang et al. suggest natural frequency perturbation (NFP) method which find the small variation of average frequency of a specific frequency band between the entry event and the exit event frequency band in the Wigner-Ville distribution (WVD)[11]. However, in previous signal processing methods, the parameters like frequency bands are manually selected to find the characteristics of the entry-exit events because those vary depending on the bearing geometry or operating conditions and the accuracy may differ accordingly, which is thus in need to more investigation into the parameter selection. To overcome the parameter-dependent results, convolutional neural network-based methods for extracting spall size-related features were studied. Guo et al. suggest hierarchical adaptive deep convolution neural network (ADCNN) to recognize the fault-pattern and fault severity evaluation[12]. However, these methods have only focused on classifying discrete fault sizes learned in the input domain. The actual spall size is continuous, and it is impossible to secure enough spall size data to sufficiently cover it with classification. And also, the entry-exit event reflected in the signal is not constant due to the uncertainty of the real spall shape or the uncertainty of rotation due to random slip. This thesis proposed a novel rolling element bearing spall size distribution estimation method using regression based deep learning. After converting the 1D acceleration signal into a time-frequency representation (TFR) through continuous wavelet transform based on complex Morlet wavelets and reducing noise that prevents finding entry-exit events

through a denoising autoencoder (DAE), the spall size is estimated using convolutional neural network with support vector regression (CNN-SVR). Since the architecture learns to find the entry-exit event-related features in the TFR during the training process, it can overcome the difficulties of manual parameter selection that varies from case to case. Then, the spall size distribution is estimated based on the average and variance of predicted values obtained from several models that have adjusted the kernel and stride size of CNN-SVR model.

1.2 Dissertation Layout

This thesis proceeds as follows. The basic concept of the spall size estimation is introduced in section 2. Section 3 describes the proposed spall size distribution estimation method and section 4 presents the validation of the proposed method using the experimental dataset. The conclusions are presented in section 5.

Chapter 2. Research Background

In this section, the method of calculating the spall size through the time interval of the entry-exit event is briefly reviewed to provide a theoretical background of the proposed spall size distribution estimation method.

2.1 Spall Size Estimation Through the Time Interval

Through the time interval between an entry event and an exit event, spall size is calculated. Figure 2-1 shows the moments when the entry event and the exit event occur. Since the entry event occurs when the rolling element center enters the spall, it is based on when the center of the rolling element is perpendicular to the entrance of the spall zone as shown in Figure 2-1(a). The exit event is based on an impulse signal that occurs when a rolling element comes into contact with the trailing edge of the spall zone as shown in Figure 2-1(b).

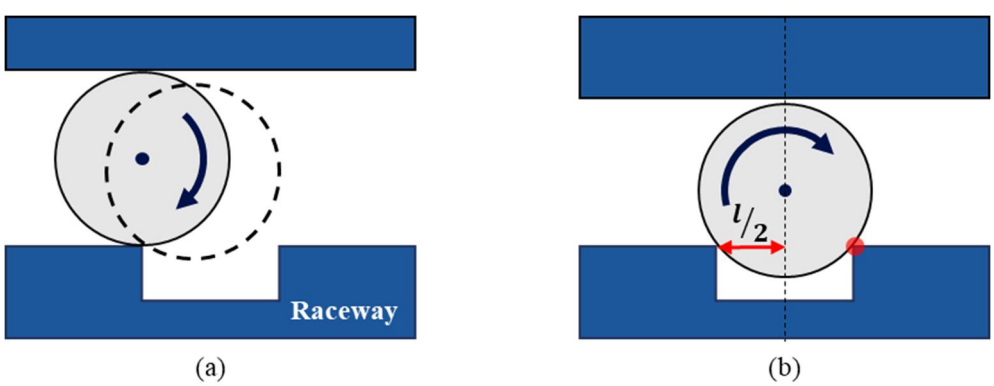

Figure 2-1. The moments of occurrence of the entry-exit event: (a) The entry event and (b) the exit event

Since the entry event and the exit event for one rolling element occur sequentially, if the time interval between the two events is known, the spall size can be estimated through the time. In particular, in the case of a small spall, which is not enough to reach the bottom of the spall zone, the distance that the rolling element passes is half of the spall size as described in Figure 2-1(b). Using this, the size of a relatively small spall is calculated based on the geometry of the corresponding bearing as follows.

$$t_o = \frac{l_o}{\pi(D_p + d)} \times \frac{1}{f_c} \quad (2\text{-}1)$$

$$t_i = \frac{l_i}{\pi(D_p - d)} \times \frac{1}{f_r - f_c} \quad (2\text{-}2)$$

where t_o, t_i are time interval of entry-exit event of outer and inner raceway, l_o, l_i are spall size of outer and inner raceway respectively, D_p is the pitch diameter, d is the ball diameter and f_r is the rotor frequency. f_c is the cage frequency

$$f_c = \frac{f_r}{2} \times \left(1 - \frac{d}{D_p}\cos\theta\right) \quad (2\text{-}3)$$

where θ is the contact angle.

Therefore, the spall size can be calculated as follows.

$$l_o, l_i = \frac{\pi f_r (D_p \pm d)(D_p \mp d\cos\theta)}{2D_p} \times \Delta t \quad (2\text{-}4)$$

where Δt is the time difference between entry and exit events.

Chapter 3. Spall size distribution estimation for REB

In this section, the proposed spall size distribution estimation for REBs using denoising autoencoder (DAE) and convolutional neural network with support vector regression (CNN-SVR) considering the uncertainty in rotational speed is discussed. The proposed spall size estimation model aims to improve the performance of estimation without additional frequency parameter setting through the regression process based on deep learning. Afterward, through the ensemble of spall size estimation models, the distribution of spall size is obtained from the signals which have the rotational uncertainty of REBs caused by the random slip of rolling elements. Figure 3-1 shows the overall procedure of the proposed method. The left side of the Figure 3-1 shows the process of training the architecture with the simulated acceleration signal of REB with the spall, and through inputting the actual acceleration signal to the trained model the spall size estimation is performed. Through the ensemble of the spall sizes, which are obtained by several models, the spall size distribution is estimated. The detail procedures are described in the following subsections.

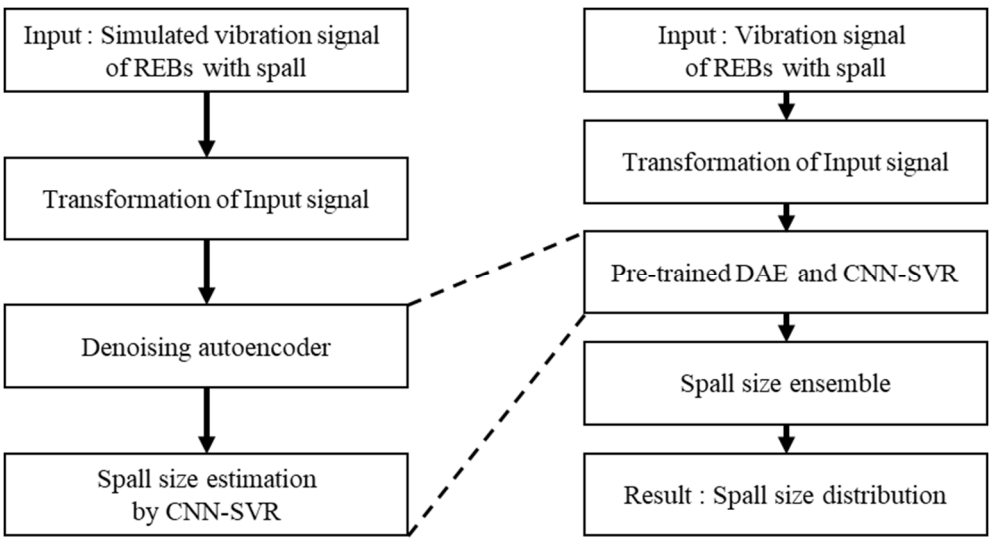

Figure 3-1. The overall procedure of the proposed spall size distribution estimation method

3.1 Transformation of Input Signal

By determining the temporal location of the entry and exit events from the acceleration signals of the REB with a spall, the spall size can be estimated by the time interval of two events. The shape of the entry event response in the time domain is the steep roll-off that occurs just after entry and the shape of the exit event response is the impact [13]. However, the entry event is masked by noises or signals of other mechanical elements used with REB, so it is difficult to accurately find the temporal location of the entry event in the time domain. In the REB acceleration signals with spall, when the rolling element passes over the spall, the entry event which is a low frequency component and the exit event which is a high frequency component are observed in sequence[7][13][14][15][16][17]. Therefore, by utilizing the frequency characteristic, a time frequency representation can be a good way to determine the temporal location of event responses. Considering the frequency characteristic and the waveform characteristics, continuous wavelet transform (CWT) is used as time-frequency representation. The acceleration signal of entry-exit event response shown in Figure 3-2(a) is transformed using CWT based on complex Morlet wavelet, as shown in Figure 3-2(b). In order to enhance the entry event signal, which is weak compared to the exit event signal, the vibration image is generated by normalizing the CWT image by time, as shown in Figure 3-2(c).

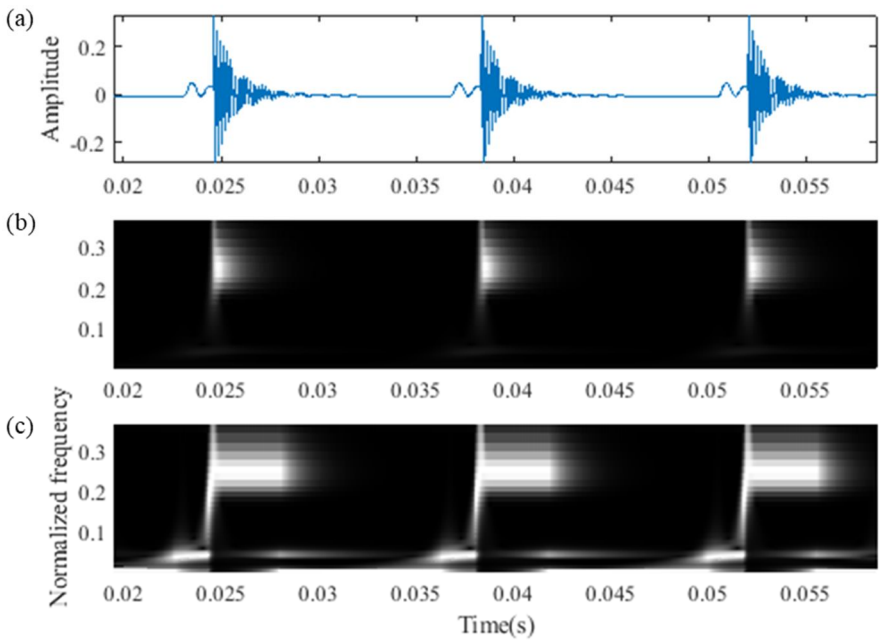

Figure 3-2. Pre-processing of raw signal: (a) the raw entry-exit event signal, (b) continuous wavelet transform based Morlet wavelet and (c) normalized continuous wavelet transform for each time

3.2 Signal Generation for Training

In order to train the deep learning models introduced in the section 3.3 and 3.4, signals similar to real ones are needed. In this thesis, analytically simulated bearing fault signals with entry-exit events were used to consider the performance of denoising and spall size estimation. The bearing fault vibration model with entry-exit event is the model in [7]. The model (y) included the impulse response (y_1), the step response (y_2) and the noise signal ($\omega(t)$). The model is expressed as

$$y_1 = e^{-\frac{t}{\tau}} \sin(2\pi f t) \quad (3\text{-}1)$$

$$y_2 = e^{-\frac{t}{3\tau}} cos(\frac{2\pi f t}{6}) + e^{-\frac{t}{5\tau}} \quad (3\text{-}2)$$

$$y = y_1 + y_2 + \omega(t) \quad (3\text{-}3)$$

where f, τ is the natural frequency (Hz) and the damping time constant (s). In order to generate similar to the actual signal, the parameters of the actual bearing signal such as natural frequency, damping time constant, spall size, rotational speed and geometric parameters are used to generate the simulated signal.

3.3 Denoising Autoencoder (DAE)

The estimation of spall size is possible through the difference between the entry event time and the exit event time. In order to estimate the correct spall size, the correct location of entry-exit event is needed. The exit event, which is an impulse response, is relatively easy to detect from noise because of its high amplitude, but the entry signal is masked by noise and its location is inaccurate. Therefore, it is necessary to effectively remove the noise of the signal. However, it is not easy to completely remove noise except for features related to spall size. Therefore, a denoising autoencoder (DAE) was used to learn the process of removing features unrelated to the entry-exit events. DAE is an autoencoder which receives a noised data and is trained to predict the denoised data of the input data[18][19][20]. Figure 3-3 shows the input image and the output image of DAE model. Since it is hard to obtain an original signal from which only noise has been removed in the real acceleration signal, the analytically simulated signals and their noise samples are used for training. The noise was added to the original signal as similar signal to noise ratio (SNR) to the actual signal. Figure 3-4 shows the encoding and decoding process of DAE. The objective function is that the difference between the noise-free data and the data with noise is minimized. Through this process, the encoder and decoder are trained to preserve the features related to the entry-exit event response while reducing unrelated features.

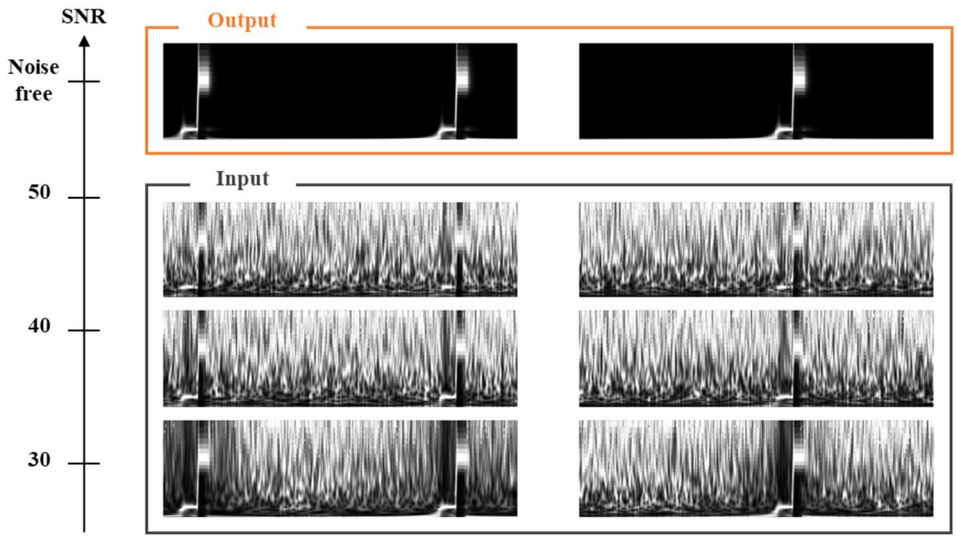

Figure 3-3. The input image and the output image of DAE model

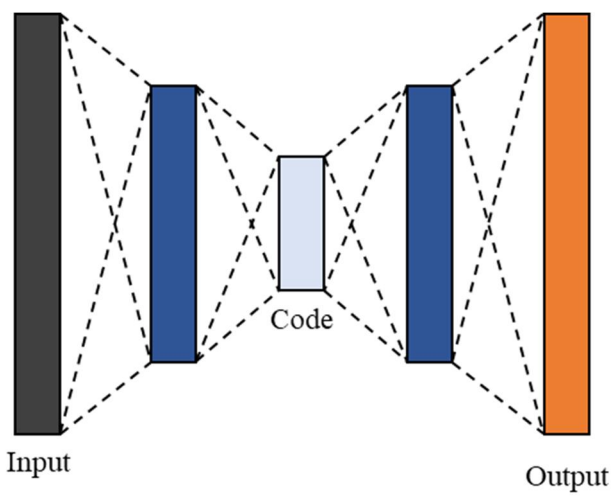

Figure 3-4. The model structure of the DAE

3.4 Spall Size Estimation Through the Time Interval

In this study, a CNN-SVR model to learn the features needed to estimate the spall size of REB is proposed. The CNN architecture learns various characteristics of the input signal through assigning each object's weights and biases and distinguishes each object accordingly and the SVR layer classify the features as the spall size[21][22][23][24][25].

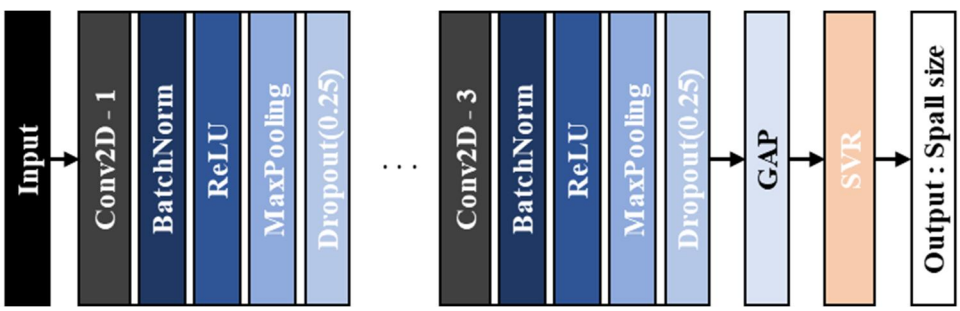

Figure 3-5. The model structure of the CNN-SVR

Figure 3-5 shows the whole model structure of the CNN-SVR. The CNN structure is composed of three feature extraction layers and global average pooling (GAP) layer. Three feature extraction layers are structurally identical, including 2D convolution layer, batch normalization, max pooling layer and dropout. The input is a binary matrix, which is the denoised signal.

The first layer is a 2D convolution layer which is designed for extracting the important local features. CNNs are classified into 1D, 2D, 3D, etc. according to the dimensionality of the input data and how the filter slides across the data. A 1D CNN is effective when derives interesting features from data and where the location of the feature within the segment is not highly relevant[26]. In this case, the entry-exit event

response is observed in a low frequency band and a high frequency band, so it is effective to use 2D CNN to distinguish occurrences in different frequency band and temporal location. The Rectified linear unit (ReLU) is used as the activation function to the convolution outputs. ReLU has a high computational efficiency and can alleviate the gradient disappearance problem[27]. The second layer is a max pooling layer connected with the outputs of previous layer. Each of the pooling windows only outputs the max value of its respective convolution layer outputs. After pooling layer, the dropout layer is stacked. Through the dropout, some elements are randomly zeros and performs a scale transformation for the non-zero parts[29][30]. In this study, the drop rate is 0.25. Therefore, the model should use the rest of the information to adapt to the target. After three feature extraction layers, GAP layer is stacked. The GAP technique could reduce the parameters and model[31][32]. By GAP, the overfitting risk of the model could further reduce. The features as output of last GAP layer are input to the SVR classifier to predict the spall size. The main target of SVR is to find an optimal hyperplane $f(x) = 0$, and $f(x)$ is defined as follows.

$$f(x) = w \cdot x + b \tag{3-4}$$

where x is the support vector, w is the weight parameter, and b is a scalar threshold. To obtain the optimal hyperplane, the positive slack variable ξ_i is used to solve the following optimization problem

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \tag{3-5}$$

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, i = 1,2,\dots,n \tag{3-6}$$

where C is a regularization parameter which penalizes the errors. At last, the

linear function $f(x)$ is transferred into a regressive function by applying a kernel function. The function for regression is derived as follows:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \exp(-\frac{\|x_i - x\|^2}{2\sigma^2}) + b \qquad (3\text{-}7)$$

where α_i and α_i^* are the Lagrange multipliers and σ is a positive real number.

3.5 Spall Size Ensemble

The optimum parameters of the architecture are determined through the spall size measured by the signal, and the size of the spall can be estimated. However, because of the rotational uncertainty caused by a random slip of rolling elements, the measured signal containing entry-exit events also has uncertainty. Due to the uncertainty of the signal, accurate spall size estimation is difficult with one optimization architecture. Therefore, by estimating the spall size through the architectures that adjusted the kernel and stride, the result values derived through the predicted values of each model. To integrate the result values, an ensemble averaging technique was applied, which is the ensemble learning methodology used to improve the prediction performance[33]. Since there are candidate models with high accuracy and low accuracy depending on the data, a weighted average is used for the result to consider that degree to the ensemble average. The distribution of predicted values for each model is derived by estimated values for each model with samples extracted from experimental data. A weighting parameter to be used for the weighted average is derived by comparing the distribution of each model with the distribution of the true value. KL divergence (D_{KL_n}) is used for weight parameter. The smaller D_{KL_n}, the greater the probability that the two distributions being compared are close. In order to consider this characteristic, the result value is derived as

$$\hat{x}_{pred} = \frac{\sum(1 - D_{KL_n})x_{pred}^{n}}{\sum(1 - D_{KL_n})} \quad (3\text{-}8)$$

where x_{pred}^{n} is the predicted value of n^{th} model and $\hat{x}_{pred}$ is the ensembled spall size of the candidate models. Through the predicted spall sizes with ensemble constructed by experimental data, a spall size distribution is constructed.

Chapter 4. Experimental Validation

In this section, the proposed spall size distribution estimation for bearing spall size is demonstrated using acceleration signal from bearing testbed data from Seoul National University (SNU). The performance of the proposed method is discussed based on the result.

4.1 Experimental Setting

To validate the proposed method, the experimental data collected from REB testbed were used. Figure 4-1 shows the REB testbed of SNU. The test bearing is located at the right end of the shaft, which is coupled by a motor shaft, and two loads are installed to apply axial and radial forces by hydraulic pump respectively. The 3-axis accelerometer is mounted vertically on the test bearing housing. The test bearings were NSK 7202A angular contact ball bearings and the artificial fault with a size of 0.602mm at the contact point was arranged as shown in Figure 4-2. The data acquisition was conducted at the rotating speed of 240, 360, 480 and 600rpm under 260kgf axial loading and 229kgf radial loading and the sampling frequency was 25.6 kHz. Each sample is 4000 sample points long and 50 samples were used for each rotating speed.

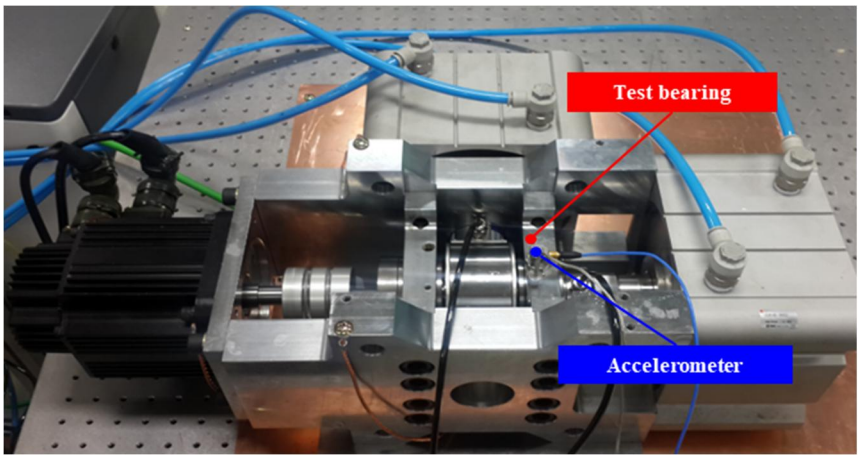

Figure 4-1. Rolling element bearing testbed of SNU

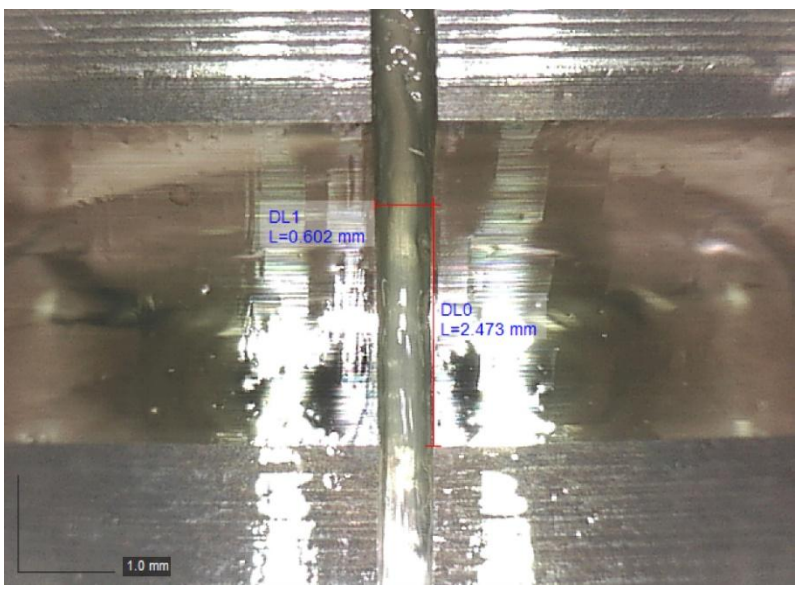

Figure 4-2. Artificial fault of test bearing with 0.602mm

4.2 Training Signal Generation

To train the DAE and CNN-SVR models introduced in section 3.3 and 3.4, analytically simulated bearing fault signals with entry-exit events were used. The bearing fault vibration model with entry-exit event is the model introduced in section 3.2. In order to be similar to the actual bearing signal, the natural frequency is set to 4000 (Hz) and the damping time constant is set to 0.001 (s) and the step response is scaled by 1/10. The signal was sampled at frequency of 25600 Hz. Since the analytically simulated signal is a signal that imitates the actual bearing signal, the time interval between the entry-exit event are set based on the actual bearing geometry. The test bearing is NSK 7202A angular contact ball bearing, and the bearing geometry parameters used to generate the signal are shown in the Table 4-1.

Table 4-1. NSK 7202A angular contact ball bearing geometry parameters

Parameter	Value
Inner ring diameter (mm)	15
Outer ring diameter (mm)	35
Ball diameter (mm)	5.9
Contact angle (°)	30
Number of balls	11

The time interval between the starting points of the step response and the impulse response is the spall size of the bearing and the interval between the starting points of impulse response is the time interval for the ball to pass through the spall. Therefore, the interval of step-impulse response was set based on BPFI, and the interval of impulse was set based on the spall size of outer race fault introduced in

Section 2, and each equation is as follows.

$$BPFI = \frac{nf_r}{2}\left(1 + \frac{d}{D}cos\,\alpha\right) \quad (4\text{-}1)$$

$$\Delta t = \frac{2D_p l_i}{\pi f_r (D_p - d)(D_p + d\cos\alpha)} \quad (4\text{-}2)$$

where n is the number of balls, f_r is the rotor frequency and Δt is the interval of step-impulse response. In this case, the rotor frequency was set to 4, 6, 8 and 10 Hz.

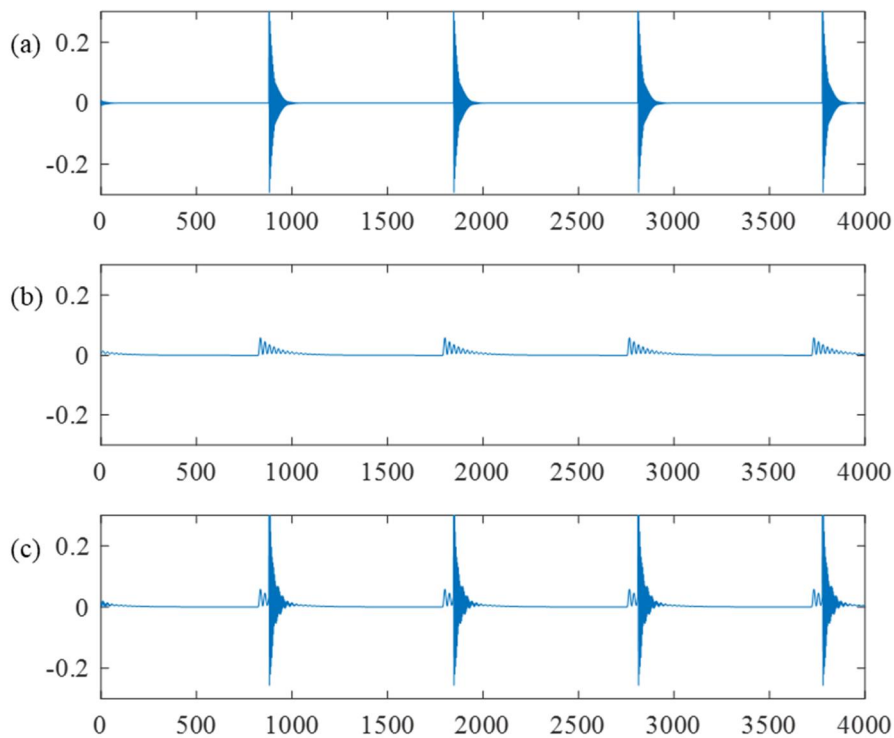

Figure 4-3. Analytically simulated entry and exit event signal: (a) the exit event signal (impulse response), (b) the entry event signal (step response) and (c) the entry-exit event signal (step-impulse response)

A step response and impulse response with the interval based on BPFI were generated as shown in Figure 4-3(a) and (b). These two responses were added by the interval of the step-impulse response to represent the bearing fault signal with outer race spall as shown in Figure 4-3(c).

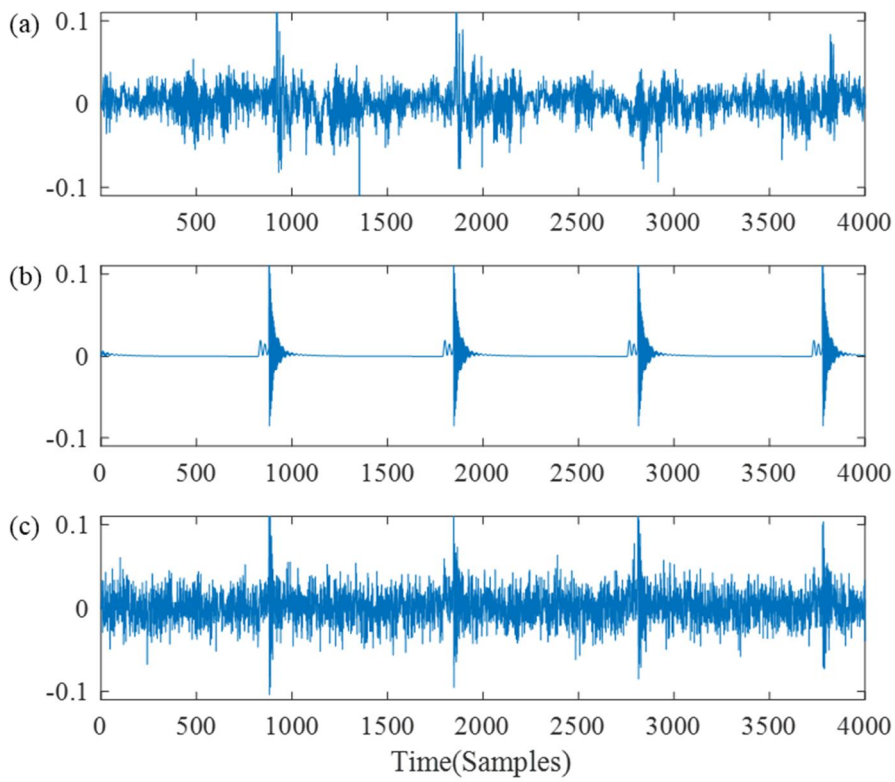

Figure 4-4. Bearing experiment data and analytically simulated data: (a) experiment data, (b) simulation data with no noise and (c) simulation data with Gaussian noise

Figure 4-4 shows the experiment data and the simulation data with no noise and with Gaussian noise. As shown in Figure 4-4(b) and (c), the simulation data were generated based on the equations used in section 3.2 and scaled based on real experiment data as shown in Figure 4-4(a).

Table 4-2. Configuration of the training and test dataset of SNU bearing testbed

Dataset	Number of samples	Spall size (mm)	Speed(RPM)
Training	100 in each case	0.3, 0.5, 0.7, 0.9	240, 360, 480, 600
Test	50 in each case	0.602	240, 360, 480, 600

The whole dataset used in training and test process is described in

Table **4-2**. Experiments were carried out at four levels of speed, and accordingly, the simulation signal was also generated based on four levels of speed. A signal with SNR of 50, 40 and 30dB of Gaussian noise added was generated to be used in DAE. In addition, the training data was set to 0.3, 0.5, 0.7 and 0.9 mm of spall size so that it does not overlap with the true value to check whether the architecture derives a value related to the spall size rather than simply classifying it through learning the training data.

4.3 Result

For comparison with the proposed method, the results of the two previous studies and the result of the proposed method without ensemble was used. The first method is a signal separation method that separates and measures the entry-exit event section on the time axis. This method is very fast and has high physical interpretability, but there is a problem due to the entry event that is easy to be masked to noise. The second method is a natural frequency perturbation method that measures using the average frequency of a selected frequency band in the Wigner-Ville spectrum, which is the time frequency representation. Since the entry-exit event is a low-high frequency event respectively, this TFR based method is valid even in a noisy environment. However, as this technique is very sensitive and sophisticated, this method has a problem in that the result is different depending on the selected frequency band.

For comparison of each result, KL divergence was used as a performance metric and the spall size distribution of the real bearing is set to mean(μ) 0.602mm and variance(σ) 0.006mm, which is 1 percent of the mean.

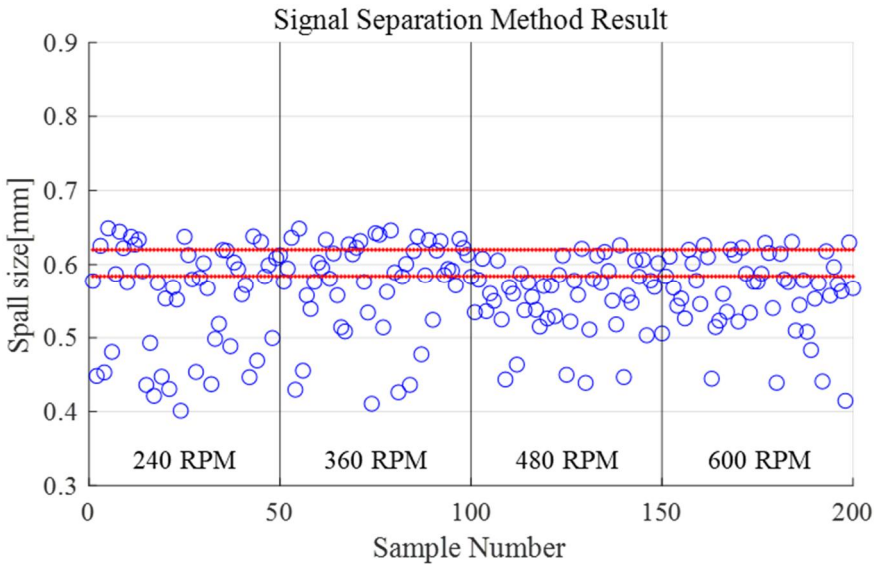

Figure 4-5. The spall size estimation result of the signal separation method at four RPM conditions

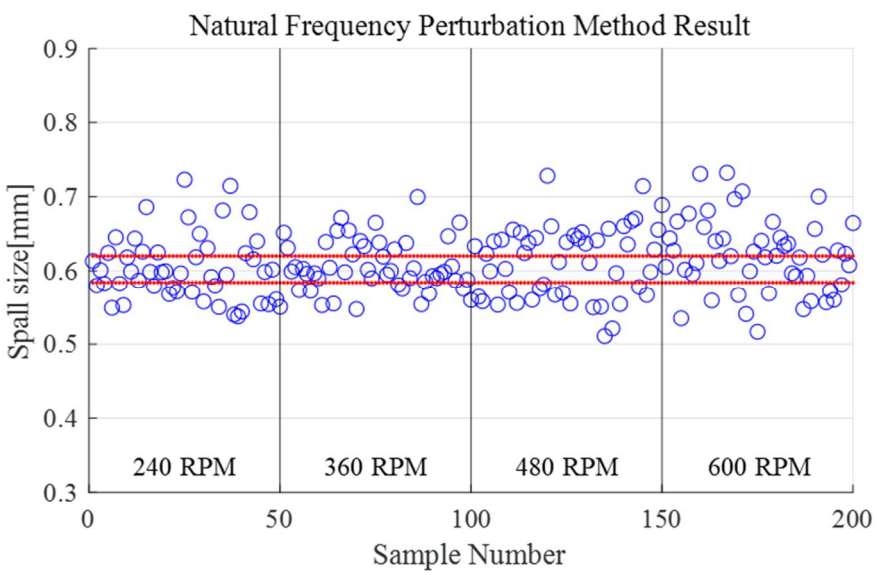

Figure 4-6. The spall size estimation result of the natural frequency perturbation method at four RPM conditions

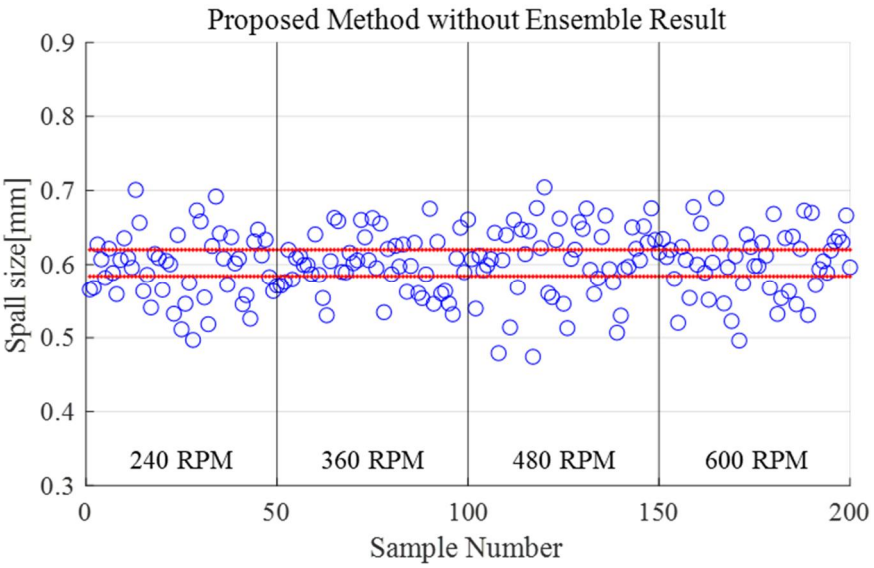

Figure 4-7. The spall size estimation result of the proposed method without ensemble at four RPM conditions

Figure 4-5,6,7 shows the spall size estimation result of the signal separation method, the natural frequency perturbation method and the proposed method without ensemble and the upper and lower limits of actual spall size set as $\mu+3\sigma$. In order to compensate for the uncertainty inherent in the signals, 20 CNN-SVR models with adjusted kernels and strides were generated and each result was derived. Through the KL divergence between the distribution of the true value made from μ and σ and the distributions of the result values of each model and the equation 3-8, the weighted parameter is calculated.

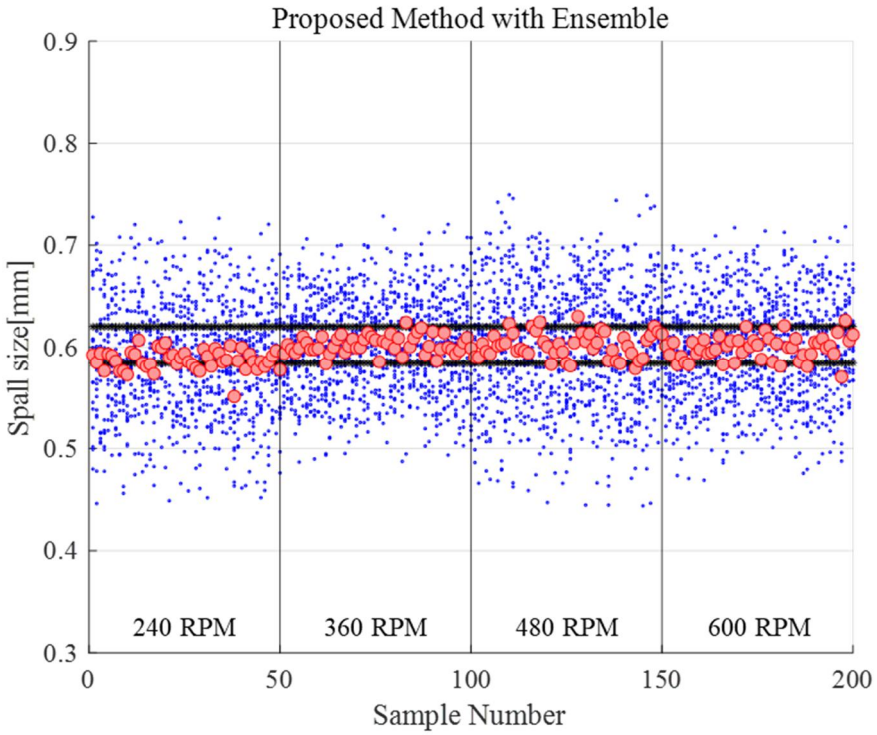

Figure 4-8. The spall size estimation results of 20 models (blue dot) and the weighted averaging result (red dot) at four rpm condition

Comparing the three methods, the result value to which the weighted ensemble average is applied is most likely within the 6-sigma range of the true value as shown in Figure 4-8. For quantitative comparison, the mean and variance were calculated by assuming a normal distribution for each result value. Table 4-3 and Table 4-4 shows the mean and variance result of four methods at four conditions. The variance of the proposed method is most similar to the true value compared to other methods.

Table 4-3. Mean and variance derived from the results at 240 and 360 rpm of the four methods.

	240 RPM		**360 RPM**	
	$\boldsymbol{\mu}$ **[mm]**	$\boldsymbol{\sigma}$ **[mm]**	$\boldsymbol{\mu}$ **[mm]**	$\boldsymbol{\sigma}$ **[mm]**
True spall size	**0.602**	**0.006**	**0.602**	**0.006**
Signal separation	0.553	0.074	0.576	0.063
Natural frequency perturbation	0.603	0.045	0.607	0.034
Proposed method w/o ensemble	0.595	0.046	0.601	0.038
Proposed method w/ ensemble	**0.587**	**0.010**	**0.602**	**0.009**

Table 4-4. Mean and variance derived from the results at 480 and 600 rpm of the four methods.

	480 RPM		**600 RPM**	
	$\boldsymbol{\mu}$ **[mm]**	$\boldsymbol{\sigma}$ **[mm]**	$\boldsymbol{\mu}$ **[mm]**	$\boldsymbol{\sigma}$ **[mm]**
True spall size	**0.602**	**0.006**	**0.602**	**0.006**
Signal separation	0.554	0.048	0.562	0.053
Natural frequency perturbation	0.613	0.049	0.620	0.050
Proposed method w/o ensemble	0.605	0.053	0.603	0.045
Proposed method w/ ensemble	**0.602**	**0.013**	**0.599**	**0.012**

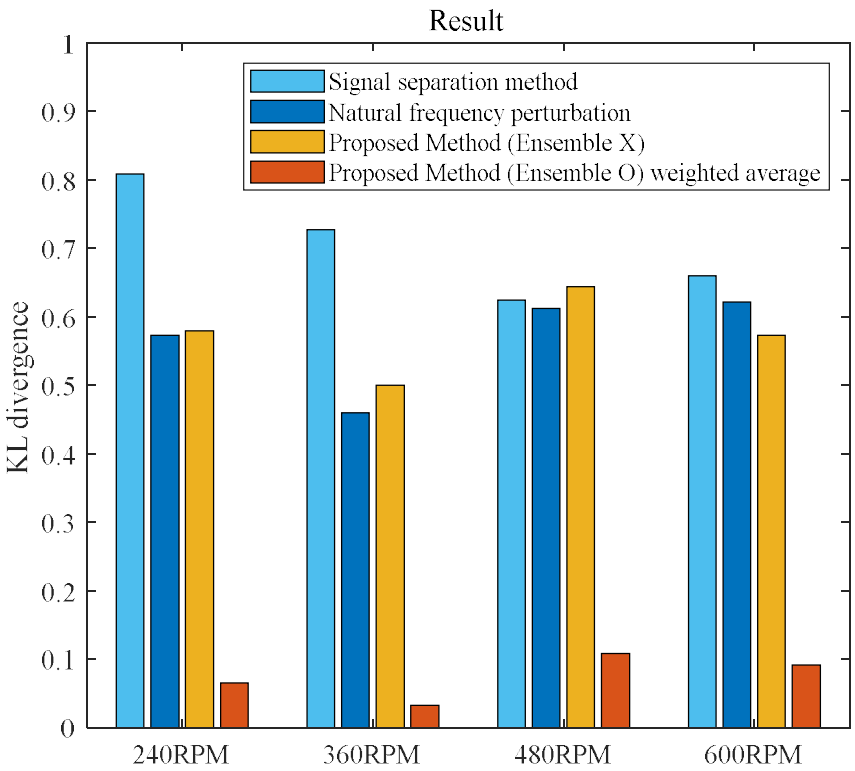

Figure 4-9. KL divergence results at four rotation speed of four methods

To quantitatively confirm the result, the KL divergence for each rotational speed was extracted. Figure 4-9 shows the KL divergence results at each speed of four methods. In the case of the signal separation method, the variance was calculated to be very large, so the KL divergence was very high, whereas in the case of NFP method, the KL divergence was relatively low because the frequency band was selected for each sample. The proposed method without ensemble estimated the spall size to a similar extent to the NFP method, and the result of adding the ensemble to this showed a very high similarity with the distribution of the actual values.

Chapter 5. Conclusions

In this study, the spall size distribution estimation for rolling element bearing using regression based deep learning was proposed.

In order to estimate the spall size through the time difference between the entry and the exit event, the time-frequency representation, CWT, was normalized by time to enhance the amplitude of the entry signal feature. The DAE is used to remove the features unrelated to the entry-exit event through a signal modeled analytically on a target signal and a noise signal with Gaussian noise. Then, the spall size of the target signal can be extracted from the signals learned through the CNN-SVR model. Through the weighted ensemble averaging, the spall size was estimated considering the rotational uncertainty due to the random slip of the balls, and through this, the spall size distribution was estimated.

To quantify the performance of the proposed method, the research described in this study employed KL divergence. The proposed method was demonstrated by experimental data from SNU bearing testbed data. By comparing the signal processing methods which use the time-frequency representation and the deep learning method without ensemble, the proposed method efficiently estimates spall size distribution.

Bibliography

[1] R. Randall, “Vibration-based condition monitoring: industrial, automotive and aerospace applications,” 2021.

[2] A. Rai and S. H. Upadhyay, “A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings,” *Tribol. Int.*, vol. 96, pp. 289–306, 2016.

[3] M. Cerrada *et al.*, “A review on data-driven fault severity assessment in rolling bearings,” *Mech. Syst. Signal Process.*, vol. 99, pp. 169–196, 2018.

[4] I. STANDARD, “INTERNATIONAL STANDARD failures — Terms , characteristics and iTeh STANDARD PREVIEW iTeh STANDARD PREVIEW,” *Iso*, vol. 2017, 2017.

[5] I. Epps, “An investigation into vibrations excited by discrete faults in rolling element bearings,” 1991.

[6] M. J. Dowling, “Application of non-stationary analysis to machinery monitoring,” *Proc. - ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 1, 1993.

[7] N. Sawalhi and R. B. Randall, “Vibration response of spalled rolling element bearings: Observations, simulations and signal processing techniques to track the spall size,” *Mech. Syst. Signal Process.*, vol. 25, no. 3, pp. 846–870, 2011.

[8] W. A. Smith, C. Hu, R. B. Randall, and Z. Peng, “Vibration-Based Spall Size Tracking in Rolling Element Bearings,” *Mech. Mach. Sci.*, vol. 21, pp. 587–597, 2015.

[9] L. Cui, N. Wu, C. Ma, and H. Wang, “Quantitative fault analysis of roller bearings based on a novel matching pursuit method with a new step-impulse

dictionary," *Mech. Syst. Signal Process.*, vol. 68–69, pp. 34–43, Feb. 2016.

[10] A. Chen and T. R. Kurfess, "Signal processing techniques for rolling element bearing spall size estimation," *Mech. Syst. Signal Process.*, vol. 117, pp. 16–32, 2019.

[11] H. Zhang, P. Borghesani, W. A. Smith, R. B. Randall, M. R. Shahriar, and Z. Peng, "Tracking the natural evolution of bearing spall size using cyclic natural frequency perturbations in vibration signals," *Mech. Syst. Signal Process.*, vol. 151, p. 107376, 2021.

[12] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Meas. J. Int. Meas. Confed.*, 2016.

[13] W. A. Smith, C. Hu, R. B. Randall, and Z. Peng, "Vibration-Based Spall Size Tracking in Rolling Element Bearings," *Mech. Mach. Sci.*, vol. 21, pp. 587–597, 2015.

[14] A. Moazen Ahmadi, C. Q. Howard, and D. Petersen, "The path of rolling elements in defective bearings: Observations, analysis and methods to estimate spall size," *J. Sound Vib.*, vol. 366, pp. 277–292, Mar. 2016.

[15] L. Cui, N. Wu, C. Ma, and H. Wang, "Quantitative fault analysis of roller bearings based on a novel matching pursuit method with a new step-impulse dictionary," *Mech. Syst. Signal Process.*, vol. 68–69, pp. 34–43, 2016.

[16] S. Zhao, L. Liang, G. Xu, J. Wang, and W. Zhang, "Quantitative diagnosis of a spall-like fault of a rolling element bearing by empirical mode decomposition and the approximate entropy method," *Mech. Syst. Signal Process.*, vol. 40, no. 1, pp. 154–177, Oct. 2013.

[17] M. A. A. Ismail and N. Sawalhi, "Vibration response characterisation and fault-size estimation of spalled ball bearings," *Insight Non-Destructive Test.*

Cond. Monit., vol. 59, no. 3, pp. 149–154, Mar. 2017.

[18] R. Thirukovalluru, S. Dixit, R. K. Sevakula, N. K. Verma, and A. Salour, "Generating feature sets for fault diagnosis using denoising stacked auto-encoder," *2016 IEEE Int. Conf. Progn. Heal. Manag. ICPHM 2016*, Aug. 2016.

[19] C. Lu, Z. Y. Wang, W. L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, vol. 130, pp. 377–388, Jan. 2017.

[20] Z. Meng, X. Zhan, J. Li, and Z. Pan, "An enhancement denoising autoencoder for rolling bearing fault diagnosis," *Measurement*, vol. 130, pp. 448–454, Dec. 2018.

[21] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," *Proc. 2017 Int. Conf. Eng. Technol. ICET 2017*, vol. 2018-Janua, pp. 1–6, 2018.

[22] L. Eren, T. Ince, and S. Kiranyaz, "A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier," *J. Signal Process. Syst.*, vol. 91, no. 2, pp. 179–189, Feb. 2019.

[23] C. Shen, D. Wang, Y. Liu, F. Kong, and P. W. Tse, "Recognition of rolling bearing fault patterns and sizes based on two-layer support vector regression machines," *Smart Struct. Syst.*, vol. 13, no. 3, pp. 1738–1991, 2014.

[24] W. You, C.-Q. Shen, X.-J. Guo, and Z.-K. Zhu, "Bearing Fault Diagnosis Using Convolution Neural Network and Support Vector Regression," *Int. Conf. Mech. Eng. Control Autom.*, 2017.

[25] W. You, C. Shen, X. Guo, X. Jiang, J. Shi, and Z. Zhu, "A hybrid technique based on convolutional neural network and support vector regression for

intelligent diagnosis of rotating machinery:," *http://dx.doi.org/10.1177/1687814017704146*, vol. 9, no. 6, p. 2017, Jun. 2017.

[26] N. Ackermann, "Introduction to 1D Convolutional Neural Networks in Keras for Time Sequences," *Goodaudience*, 2018.

[27] F. Godin, J. Degrave, J. Dambre, and W. De Neve, "Dual Rectified Linear Units (DReLUs): A replacement for tanh activation functions in Quasi-Recurrent Neural Networks," *Pattern Recognit. Lett.*, vol. 116, pp. 8–14, Dec. 2018.

[28] G. Lin and W. Shen, "Research on convolutional neural network based on improved Relu piecewise activation function," *Procedia Comput. Sci.*, vol. 131, pp. 977–984, 2018.

[29] A. Achille and S. Soatto, "Information Dropout: Learning Optimal Representations Through Noisy Computation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2897–2905, Dec. 2018.

[30] H. Wu and X. Gu, "Towards dropout training for convolutional neural networks," *Neural Networks*, vol. 71, pp. 1–10, Nov. 2015.

[31] W. Gong, H. Chen, Z. Zhang, M. Zhang, and H. Gao, "A Data-Driven-Based Fault Diagnosis Approach for Electrical Power DC-DC Inverter by Using Modified Convolutional Neural Network with Global Average Pooling and 2-D Feature Image," *IEEE Access*, vol. 8, pp. 73677–73697, 2020.

[32] M. Lin, Q. Chen, and S. Yan, "Network In Network," *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc.*, Dec. 2013.

[33] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, pp. 1–18, 2018.

국문 초록

구름요소 베어링 진단을 위한 딥러닝 기반 스폴 크기 분포 추정 연구

서울대학교 공과대학
기계공학부 대학원
정 화 용

구름 접촉 피로로 인한 스폴은 구름 요소 베어링 파손의 가장 일반적인 원인이며 스폴 크기 추정은 심각도를 추정하는 좋은 방법이 될 수 있다. 기존 연구에서는 구름 요소가 스폴 영역을 지나가는 과정에서, 진입할 때 저주파 단계 응답이 나타나고, 이탈할 때 고주파의 충격 응답이 나타난다고 알려져 있다. 진입이벤트 신호는 이탈이벤트 신호 및 노이즈에 비해 상대적으로 약하기 때문에 지금까지의 연구에서는 노이즈 감소, 진입이벤트 특성인자 강화 등의 다양한 신호처리 기술을 이용하여 진입이벤트의 시간적 위치를 추정하고자 하였다. 그러나 신호처리에서 진입이벤트의 특성을 찾기 위한 매개변수 선택은 베어링 형상이나 작동 조건 등에 따라 다르며, 선택이 경험적이므로 정확도가 경우에 따라 다를 수 있다. 또한 신호에 반영된 스폴의 크기도 실제 스폴의 일정하지 않은 모양에 의한 불확실성과 베어링 구름요소의 임의 미끄러짐으로 인한 회전의 불확실성을 가지고 있다. 이러한 어려움을 극복하기 위해

본 연구에서는 딥러닝 기반 접근 방식을 제안한다. 테스트 데이터와 유사한 형상 및 작동 조건인 해석적 시뮬레이션 신호를 통해 학습된 제안모델은 매개변수의 수동적 선택 없이 스폴의 크기를 추정한다. 여러 커널과 스트라이드가 선택되어 만들어진 여러 훈련모델에서 얻은 추정값을 통해 평균과 분산을 구하여 파편 크기 분포를 추정한다. 제안된 모델은 고장을 인가한 베어링을 통해 얻어진 실험 데이터로 검증한다. 성능 분석 결과는 제안된 접근 방식이 효과적임을 나타낸다.

주요어 : 스폴 크기 추정
구름 요소 베어링
고장 진단
디노이징 오토인코더
합성곱 신경망
앙상블학습

학 번 : 2018-27800