Noname manuscript No.
(will be inserted by the editor)

An One-Stage Ensemble Framework based on Convolutional Autoencoder for Remaining Useful Life Estimation

Yong-Keun Park · Min-Kyung Kim · Jumyung Um*

Received: date / Accepted: date

Abstract The research on predictive maintenance of rotating machines, the most important element in manufacturing facilities, has been very active. The widespread availability of smart factory solutions has led to improved data collection from machines and processes and is able to provide key information. For our purpose, the collected information enables the maintenance system to predict the remaining useful life using deep learning models. The introduction of multi-layer perceptron of signal processing originating from bearings, in time series data, has been discussed in many publications. However, estimating accuracy for the remaining useful life is determined by the selection of the feature domain and the concatenation network model. Herein, we introduce a convolutional Autoencoder based on multi-domain ensemble learning in order to include various feature domains and a concatenation network operated by latent space into a single neural network. The performance of the proposed model is evaluated by using a simple health indicator and a PRONOSTIA dataset and compared with a simple concatenation model, 2-stage Autoencoder, and a recurrent neural network.

Keywords Remaining useful life · Convolutional auto-encoder · Ensemble learning · Prognostics health management

1 introduction

The field of PHM(Prognostics and Health Management) was initiated from electronics area in 2006 [1] and extended to the are of rotary machinery in 2014 [2]. It is intended to provide the increase of productivity in machine

Y-K. Park, M-K Kim and J. Um
Kyung Hee University, 1732 Deogyeong-daero, Yongin-si, Republic of Korea
Tel.: +82-31-201-3695
Fax: +82-31-203-4004
E-mail: jayum@khu.ac.kr

maintenance. Since 2011, Industry 4.0 widely spreads and was initiative to build the infrastructure able to collect data from device, machine and whole system. And an increasing amount of literature has been published in PHM over a diverse methods from machine learning to deep learning.

1.1 prognostics and health management

At the PHM stage, data is collected using sensors, and after determining the health of the system based on the collected data, an abnormal situation is diagnosed and RUL (Remaining Useful Life) is predicted. In addition, HI (Health Indicator) extraction, which is an indicator of health for evaluating machine health, is an indispensable process. Finally, the time of machine failure can be predicted via HI. It may result in the completion of PdM technology that can be serviced only when needed. In addition, based on this, the optimum maintenance cycle and range for formulating a maintenance strategy can be determined, so there is a cost reduction effect[5].

For predictive maintenance of machine, bearing is the main component of analysis. There are many rotating part in machine. Bearings account for about 40% of machine failures[6] and are the most breakdown component in failure.

For these reasons, among the various component of machine, bearing remaining useful life prediction is used in the main predictive maintenance method. Bearing remaining useful life prediction is mainly analyzed via vibration signals and sound[7], and various signal processing technologies such as noise reduction and fault frequency measurement are used.

There are several methods to extract features in various areas; however, the use of a single domain is problematic. For example, if only the time domain model is used, a prediction model skips learning the crack signal occurred in the frequency spectrum; thus, the model can miss the information that the features have in each domain, and even though some models and machines perform well, the problem resides in that the pre-processing method depends on the data specific to each domain. Therefore, many studies have suggested features created from various domains of time, frequency, and time-frequency in order to complement the shortcomings of aforementioned methods consisting of extracting a single domain feature. However, despite these attempts, there exist drawbacks, since the features of the three domains are lumped into one data set, i.e simple concatenate, that is used as the input of the model. This method is insufficient to efficiently extract the information included in each domain, and to conduct as a solution for multi-domain features extraction.

Selecting the best feature data set or using latent space of AE (Auto Encoder) were the typical methods to achieve multi-domain feature extraction. Both methods, however, have some drawbacks. Using a particular feature makes a learning model rely on specific metrics. It diminishes the benefits to build a data set from multiple domains. For example, it is worthwhile to add a specific feature capturing the critical moments of the breakage, even though giving the model a negative influence in the most period during the

learning process. Feature selection ignores features showing the correlation in only a certain moment.

To eliminate the drawbacks of feature selection, several studies have adopted latent space to reduce the dimension using deep learning models, in particular AE , in feature extraction. However, feature extraction through AE has a complexity problem arising from the need to learn two different prediction models: the first AE for learning and the second latent space and RUL for predicting HI. Learning the two models has the disadvantage of increasing the complexity of the model structure, resulting in that efficient learning of the latent space is not performed. The reason for this is that after learning AE, features are extracted using the latent space. The extracted features are then used to train the RUL prediction model. The disadvantage of this method is that HI, which is the final output value of the RUL prediction model, is learned without any interaction with the features extracted through the latent space.

In this paper, the authors suggest the method using an all-domain data set in a one-stage learning model including latent spaces of AE and RUL model to predict HI. The remainder of this paper is structured as follows: section 2 surveys the state-of-the-art of relevant research. section 3 describes the proposed model architecture. Section 4 compares the proposed model with relevant models of HI prediction. Section 5 provides the discussion and conclusion.

2 Literature survey

Sensor data driven approach is required for pre-processing of sensor signal. Signal processing techniques are used to analyze vibration signals. Preprocessing is performed on 3 main domain. There are 3 types: (1) time domain feature, (2) frequency domain feature, and (3) time-frequency domain feature. (1) In the case of time domain features Various statistics have been widely used to efficiently capture bearing failure signals. Typically, RMS (Root Mean Square), Kurtosis, etc. have been used. These statistics describe the overall failure process of bearing. However, there is a drawback that it does not describe the detailed failure steps. (2) Frequency domain features are mainly analyzed by transforming time domain signals such as FT (Fourier Transformation) and Hilbert-Huang transformation to the frequency domain. The advantage of these features is that the technology is known to efficiently perform the steps of initial failure and final failure. However, it has the disadvantage of not describing the procedure for intermediate failures. (3) Time-Frequency domain features are often STFT (Short Time Fourier Transform) and WT (Wavelet Transformation). These transformation features are known as have few errors without loss of information[11].

After the preprocessing is done, selecting meaningful features is performed. Such a method is called feature selection. Feature selection select more efficient feature for obtaining bearing failure sign and removing unwanted features. in this process, various metrics are used, that are monotonicity, trendability, robustness, etc[12]. Features are selected based on these metrics. The process for

selecting sensitive features includes not only feature selection, but also feature extraction. Feature extraction is a method of creating new features by using existing features. PCA(Principle Component Analysis), ICA(Independent Component Analysis), Partial Least Square, SOM(Self-Organizing Map), etc. are used[13]. In addition to this, a feature extraction method through deep learning is also used, using a method in which a model or neural network extracts its own function using CNN(Convolutional Neural Network), or using AE(AutoEncoder). AE is one of the feature extraction methods that has attracted a great deal of attention by utilizing the latent space.

[12] [13] [14]

RUL prediction has been extensively studied, but two general methods have been widely used. The first method is to predict the change of breakage through a physical model[8], and the second is to predict the degradation based on sensor data[9]. The physical model is practically difficult to implement a complex actual facility operating environment. In other words, expressing all situations with few failure mechanisms is bound to be limited. For this reason, machine learning techniques that learn from sensor data based on a black box model are widely applied[10].

2.1 Domain Selection

In many studies, to analyze the bearing vibration signal, various features have been used in 3 domain which are time domain, frequency domain, and time-frequency domains.

Using time domain as input features is to apply a set of statistics factors extracting from time-series signal data. Fundamentally RMS (Root Mean Square) have been used as bearing diagnosis and HI [16]. Adding the window size in RMS is suggested to depict the changes along the time [17]. Among the statistics used as features in the time domain[15], a statistic called waveform entropy was presented. In [18], after decomposing signals into specific levels using Wavelet Packet Decomposition (WPD), 14 statistics were extracted for each decomposed signal level. In addition, many studies have been conducted using the original signal[19]. The window size is adjusted and the original signal is used as an input. In addition, there have been attempts to find a meaningful information only the original signal through CNN[20].

[21] constructed input data by using the frequency signal applying a moving average after applying FT. In addition, [22] and [15] research using FT as input data.

Finally, there are studies using the time-frequency domain. There are also studies in which an input shape similar to an image in a two-dimensional format was created using STFT (Short Time Fourier Transformation) and learned through CNN[23]. In addition, [24] imaged each signal using WT and then extracted HI using a CNN network.

There are studies using a multi domain input. [25], [11], [26] make features using statistics as a time domain feature, FT as a frequency domain feature,

and WT as a time frequency domain feature. [27] also make feature in 3 domains, reduces the dimension through RBM (Restricted Boltzmann machine), and predicts HI through GRU (Gated Recurrent Units). In addition, [28] that do not use all 3 domains and use statistics and raw signals in some cases, and [29] use an RVM (Relevance vector machine) to construct a model. [30] predicted HI using statistics and frequency domain feature.

2.2 Feature selection and Feature extraction

After collecting a feature in each area, both feature selection and feature extraction remove unnecessary signals and remain only useful information. Various metrics such as monotonicity, trendability, and robustness are used for feature selection. In [15], only trendability among features is applied to machine learning. In other machine learning techniques, [26] selected three metrics, trendability, monotonicity, and the average of the two metrics, while [18] used Mahalanobis distance of features. [31] highlighted that there have also been studies to select features by using information theory metric. These studies are to use statistical techniques as feature selection. Sensitive features using the chi-square test[32] and F-test[33] are applied to feature selection.

Following feature selection, there are many studies in which AE is performed as a feature extraction. Lin and Tao extracted features using a latent space after ensembles a number of AEs, and used it as an input to a model that predicts RUL [34]. Similar studies using three domains as feature sets conducted, But Ren et al. used latent space and the remaining set of domain features as inputs to the RUL prediction model after feature extraction via AE only in the time domain [11]. Xu et al also applied AE model to only prediction of final output value of HI [21]. AE is a deep learning model with the same input and output, but there is a difference in that the input data is not restored. Similarly, there is a PHM study that does not restore input data [35], and the difference is that a network is designed for bearing failure classification from the final output value. Some studies tried to conduct both feature selection and feature extraction simultaneously. Hu et al. and She et al. performed feature extraction through RBM and then selected features through HI metric [30][25].

2.3 Deep learning model of remaining useful life

After extracting the feature for bearing failure from the vibration signal in various domain, the remaining useful life is predicted via a deep learning model. Deep running models typcially used MLP(Multi Layer Perceptron), and uses RNN(Recurrent Neural Network) and CNN [14]. The remaining useful life was efficiently predicted through the features extracted in the above. After demonstrating the efficiency of the latent space using AE, AE was used to reduce the dimensions and perform feature extraction. The features extracted

in this way were used as inputs to the model for predicting the remaining useful life, and created a bearing prediction model with even higher performance [9].

2.4 Summary and Opportunities

As the survey of previous researches, many studies have used specific domain to extract features. There are cases where a single domain is used as a feature, and other studies used two or more domain as a feature. However, A common point of both cases is that feature extraction was simply made into one data set and used as an input to the model. The problem of using only one domain as a feature is missing the characteristics of remaining features. This issue arises even when more than one domain is used as a feature. It is simply concatenate of one data set and used to input the model. Simple concatenation is insufficient to learn all domains efficiently.

From previous literature, AE secures the better performance of feature extraction and of feature selection. There have been various researches on feature selection and extraction, but accuracy of learning models fluctuated, because feature selection goes through a process of selecting features based on specific metrics. Feature extraction via AE shows the high accuracy than using other dimension reduction methods. However, learning two models at the same time, feature extraction via AE and RUL prediction, increase model complexity unnecessarily. As a solution to this, there have also been studies predicting HI directly through AE. Even in these cases, the difference between inputs and outputs appear a problem that the latent space cannot be trained efficiently. the study of constructing the two models into one model are insufficient. In this work, we present a model consisting of one model, along with latent spatial layer learning of AE. The conclusions we can draw after reviewing the literature are:

1. Feature sources need to be rich and diverse. Estimating remaining useful lifetime require for multiple aspects of bearing sensor signals. As existing literature describe, time-series data, frequency data, and time-frequency data are employed for feature extraction because bearing signals usually are vibration, current, and acoustic noise.

2. Feature selection is required for reducing abnormal effects from whole dataset. The disturbance of machine learning is caused by the series of data not to follow the learning direction of whole dataset. It is noise or event happened by very specific accident. Selecting features is critical decision to reduce the effect of noise like this. It is inevitable to ignore useful data too. The abandon of certain data need to logical reason in order to keep consistence of estimating RUL.

3. Feature synthesis improves the possibility of finding hidden correlation of raw data. Both mathematical approach and data driven approach are used for evaluating correlations and reducing data dimension.

From these conclusions it appear that there is opportunity to motivate new predictive maintenance algorithm of rotary parts.

3 Feature Ensemble AutoEncoder Architecture

Motivated by the opportunities mentioned in previous section, then, the improvement of feature extraction, and selection are viewed as the way of enhancing the performance of RUL prediction. In this paper, it is proposed for comprehensive data flow turned into an architecture incorporating ensemble learning with FEAE(Feature Ensemble AutoEncoder) and is depicted in Figure 1. The proposed model framework is design for covering with following aspects

1. Wavelet Packet Decomposition to distinguish low and high frequency in time-frequency domain.
2. Applying AutoEncoder to Multi-domain approach
3. Reducing dimensions of multi-domain by using latent space generated by AutoEncoder
4. 1-stage model with combined loss function is applied to enhance the learning rate of whole model.

Signal processing technique is used to remove noise and redundant information of signal. Then after applying signal processing technique, extract various domain features that time, frequency, and raw signal source, it is used as input data of the FEAE model. The FEAE model consists of two parts. First, FEAE learn latent space through each AE. At the same time, each latent space is coming through input of the RUL model, and finally HI is predicted.

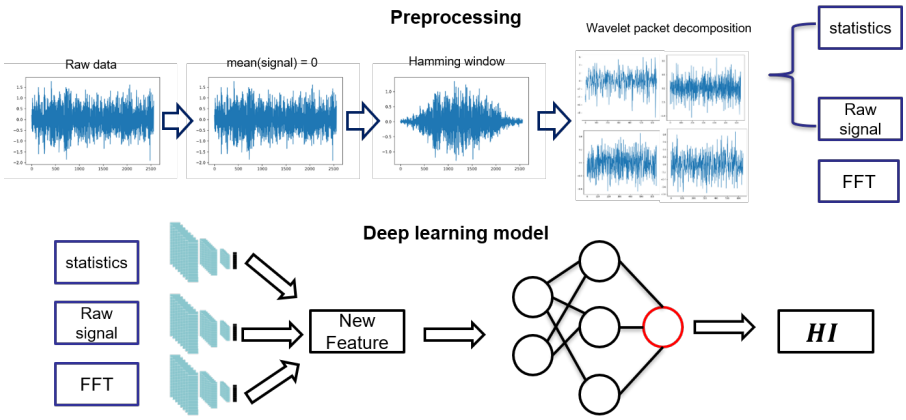

Fig. 1 Procedure of proposed framework for estimating RUL

3.1 Signal processing of Wavelet Packet Decomposition

Since bearing vibration signal has rich information such as trouble, faulty of machine, accurate vibration analysis is essential process. First, Direct Component(DC) value remove from the signal purpose of right transformation such as fourier and wavelet. Second, to remove noise in signal apply window function because of edge of signal has many uncertainty.

Wavelet Packet Decomposition(WPD) is a time-frequency analysis method that is widely used in the field of signal processing. We apply WPD to preprocessing of proposed procedure in order to improve the resolution of complex signal data. In the case of WPD, the signal is decomposed into coefficients (low frequency components) and details (high frequency components) at the first level. This can be thought of as the low-frequency and high-frequency pass filter components of the signal[36].

Zhang, Z., Wang, Y., Wang, K. (2013). Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network. Journal of Intelligent Manufacturing, 24(6), 1213-1227.

The main difference between the WT(Wavelet Transform) and WPD is the way the signal is more decomposed after the first level. Simply put, WPD are WT that pass more filters in signal processing. WT decomposes only low-frequency components at subsequent levels, while WPD decomposes low-frequency and high-frequency components at each level[37]. Therefore, WPD provides better resolution in various areas where the signal contains high frequency information.

The equation $W\left(p,q\right)$ presented below is an equation for WT. $x\left(t\right)$ is the signal to apply WT, $\psi^{*}(t)$ means a certain wavelet (mother wavelet), and ψ^{*} means complex conjugate. To do. Further, p plays a role of reducing or increasing a certain wavelet as a scaling parameter, and q means a variable that moves a certain wavelet as a movement parameter. The formula DWT(Discrete Wavelet Transformation) (i,j) can be obtained from the dioxide of $W\left(p,q\right)$. p and q are replaced by 2^{i}, 2^{j}. Repeated decomposition can be used to obtain low frequency (cA_1) and high frequency (cD_1) band signals figure 2

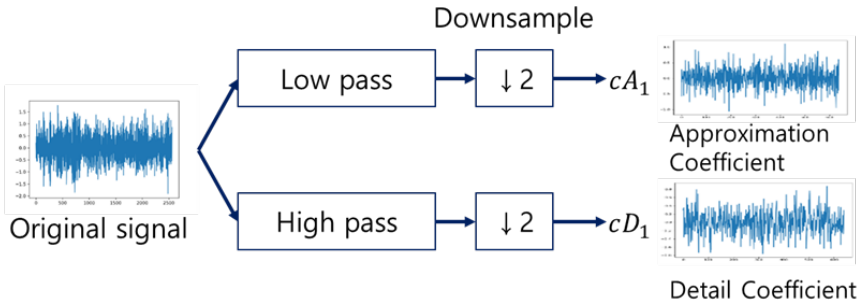

Fig. 2 Procedure of Discrete Wavelet Transformation

When the original signal is decomposed into level 2 via the WPD process, it can be decomposed into a total of 4 signals as shown in figure 3. WT is said to be a nonstationary signal that works as a frequency filter and is effective in reducing noise [38], [39]. The decomposed signal contains low-frequency to high-frequency features. Next, in order to execute the preprocessing of the input data, the four decomposition signals and the original signal are created as one data set, and then the preprocessing is executed.

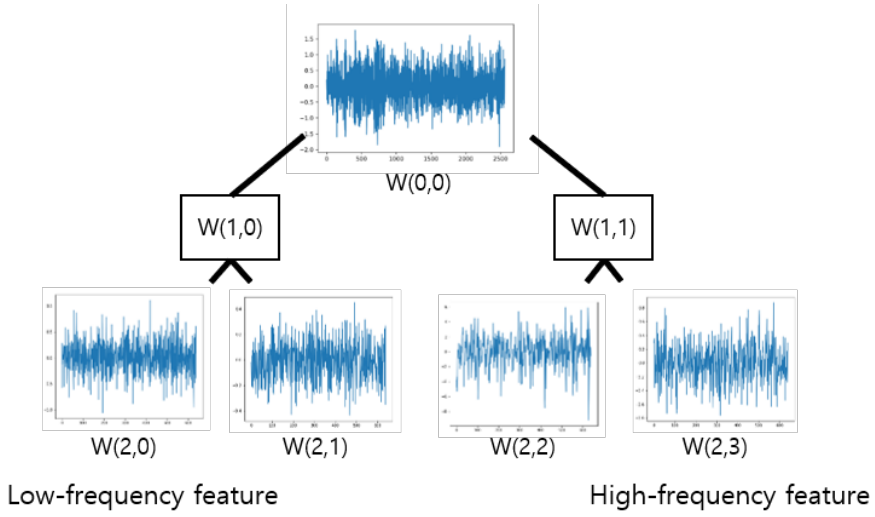

Fig. 3 Decomposition result of WPD up to Level 2

In particular, the resolution of WPD helps to detect progressive increase of specific vibration caused by the growth of cracks in rotating components. Figure 4 shows the advantage of WPD in estimating RUL. The input signal data is to made by combining narrow low-frequency signal, decreasing signal, increasing signal, wide high-frequency, and random noise signal in all different frequencies. The result of WPD keeps the trends of each signal even though the noise signal is mixed. The accuracy of WPD resolution is evaluated with the result of FFT filter and DC removing filter. Orange peaks cover with blue peaks of original signal.

3.2 Preprocessing of Multi-domain dataset

In this section, we will turn our focus to the way to extract feature sets. From raw signal, pre-processing is required to extract bearing degradation signal efficiently. Typically feature sets of bearing data can be categorized into Signal, Statistics, Frequency. In the case of time-domain basis feature set, several traditional features were introduced for predicting the failures of bearings such

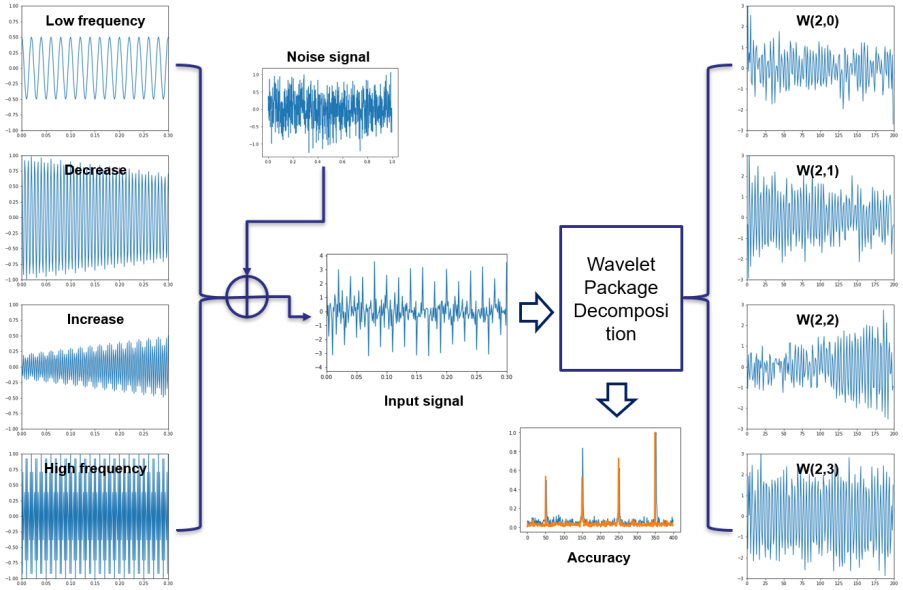

Fig. 4 Resolution result of different mixed signals by using Level 2 WPD

as RMS, kurtosis, skewness, etc. In the case of frequency-domain basis feature set, Fourier transform that transforms time domain vibration data into a frequency spectrum is used to extract frequency domain information. In the case of time-frequency domain basis feature set, wavelet analysis is used to analyze signal as different resolution. In this paper uses 3 domain features that are below discussed

1. Signal: Original signal & wavelet signal are used to signal domain feature
2. Statistics: Statistics of original signal & wavelet signal are used to statistics domain feature
3. Frequency: Unlike above mentioned features, Frequency feature only work original signal.

3.2.1 Signal feature set

The basic signal is an input type that is widely used in various studies applying deep learning. This has the advantage of not requiring preprocessing. Also, in the case of basic signals, it is used in the most common form to which the basic signal processing algorithms and WPD described in Section 3.2 are applied. As shown in Figure 5, in the case of the final data set, it is used as one data set to which the original signal and the wavelet are applied.

3.2.2 Statistics feature set

Extract the statistic of the converted signal and the original signal using WPD. Statistics often used in RUL prediction are RMS, Kurtosis, Peak-to-peak, Mar-

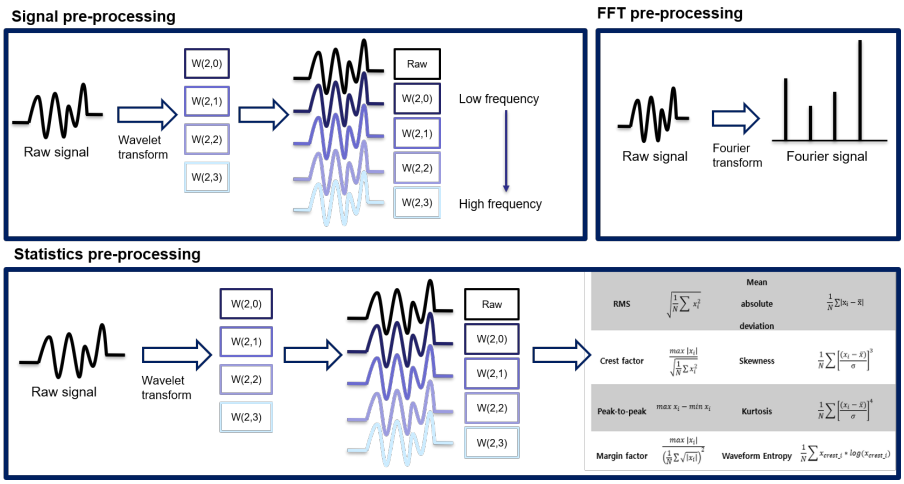

Fig. 5 Pre-processing of each signal source

gin factor, Mean absolute deviation, Skewness, Kurtosis, and Waveform entropy. A total of 8 statistics are used. Since eight statistics of the wavelet-transformed signal and the original signal are extracted, a total of 40 types of statistics are extracted. Therefore, it is possible to efficiently extract various information possessed by each frequency signal. The overall procedure for the Statistics feature set is shown in Figure 5.

The proposed model uses the easy-to-obtain frequency domain signals as the input of the deep learning model basically. Time-domain features are correlated with the overall trend for the bearing faulty and are the fundamental inputs of deep learning model for machine degradation. But not only that, but also because of the nature of raw data that is not found with frequency filters, primary statistics of raw data is selected as first input of proposed model. It is also includes statistical figures directly obtained from raw data and a method for collecting raw data characteristics, unlike in the frequency domain.

3.2.3 Frequency feature set

FT has the advantage of being able to move signals in the time domain to the frequency domain. In addition, various frequencies of the signal can be decomposed by frequency via a periodic function to confirm the characteristics in an intuitive understanding of FT. The following equation is a Fourier transform equation that decomposes a signal via a sine periodic function. In the case of Fourier transform, it plays a role of converting the signal in the time domain into the frequency domain. However, since the signal obtained via WPD is already a signal decomposed for each frequency, the same information will be obtained when FT is applied. For these reasons, as shown in Figure 5, WPD was not applied to Frequency, which is a frequency domain feature set, and only FT was applied to the original signal.

3.3 Ensemble learning of latent space generated by AutoEncoders

In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. Unlike a statistical ensemble in statistical mechanics, which is usually infinite, a machine learning ensemble consists of only a concrete finite set of alternative models, but typically allows for much more flexible structure to exist among those alternatives . An ensemble model method is a machine learning process to obtain better prediction performance by strategically combining multiple learning algorithms . There are three primary advantages brought by ensemble methods . 1) The first one is called statistical reason which is related to lack of sufficient data to properly represent the data distribution. Without sufficient data, many hypotheses which give the same training accuracy may be chosen as the learning algorithm. Ensemble methods can thus reduce the risk of selecting the wrong model by aggregating all these candidate models. 2) The second is computational reason. Many learning algorithms, such as decision tree and neural network, work by performing some form of local search. These methods can frequently result in locally optimal solutions. Ensemble methods show their advantages in this scenario by running many local search from different starting points. 3) The last reason is representational. In most cases, the true function f cannot be represented by any single hypothesis H. However, the function can be better approximated by a weighted sum of several hypotheses.

3.4 1-stage learning model for RUL prediction

Finally, all the methods described in the methodology are used to provide a new model, FEAE. First, proceed with the ensemble of input data. In order to efficiently compress and extract the information that each input data has, the structure shown in figure 6 is used. This structure is a model based on three AEs, and each model uses a feature set of three regions as input values. Each feature set is the three input data of Signal, Statistics, and Frequency described above. Also, the latent space of each AE is connected to the RUL model that predicts HI. The AE and RUL models consist of one stage, that is, one model.

There are two main models proposed, and the first is AE that learns latent space. The second part is a model that predicts HI using the latent space as an input to the RUL model. FEAE will eventually use the feature set of the three areas Signal, Statistics, and Frequency as input. In addition, the final output has a total of 4 outputs with 3 AE restoration results for each feature set and 1 RUL model. The advantage of the structure of these models is that the latent space of AE is efficiently trained by two elements. First, the latent space is trained through the restoration loss function of AE. Second, each latent space is used as an input to the RUL model and predicts HI. There is an advantage

that the latent space can be learned via MSE (Mean Square Error), which is the loss function of this RUL model.

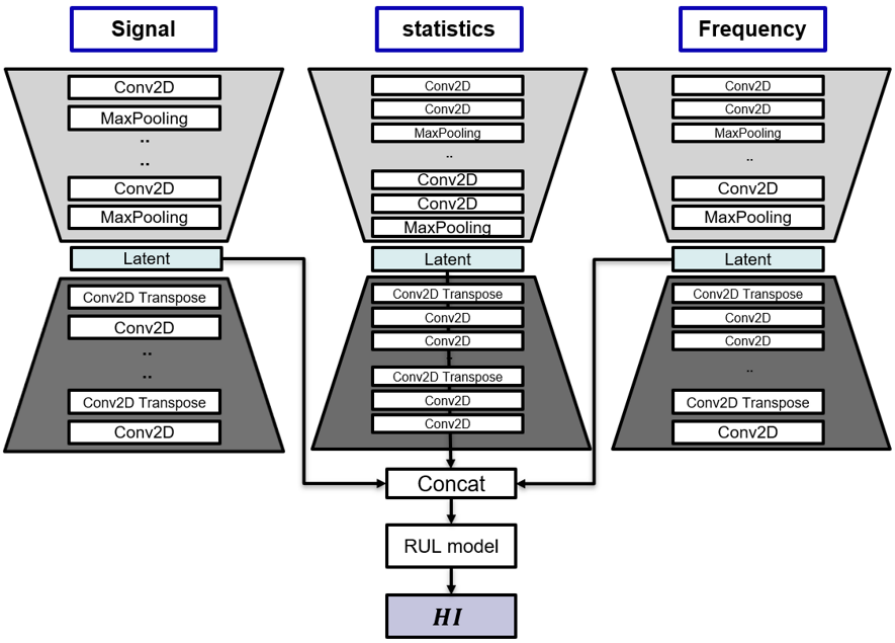

Fig. 6 Network model of 1-stage feature ensemble autoencoder for estimating RUL

3.5 Health indicator of remaining useful life

There are many ways in which HI can be designed. However, HI, which predicts the failure time of another machine, is generally highly volatile, making it difficult to set a failure threshold. In general, even in the case of bearings, the threshold value is determined experimentally because the HI value is different for each work and data. To dispel these ambiguities, this paper also uses common HI design methods. The HI value in the initial state is set to 0%, and when a complete failure occurs, the HI value is set to 100% to present the HI value that can be applied in any situation. For example, if the total drive time is 2800s and the current time is 1400s, the current HI is set to 0.5 fiture 7.

$$(x_t, y_t), where\ t \in T, x_t \in F,\ (F = Feature\ set) \tag{1}$$

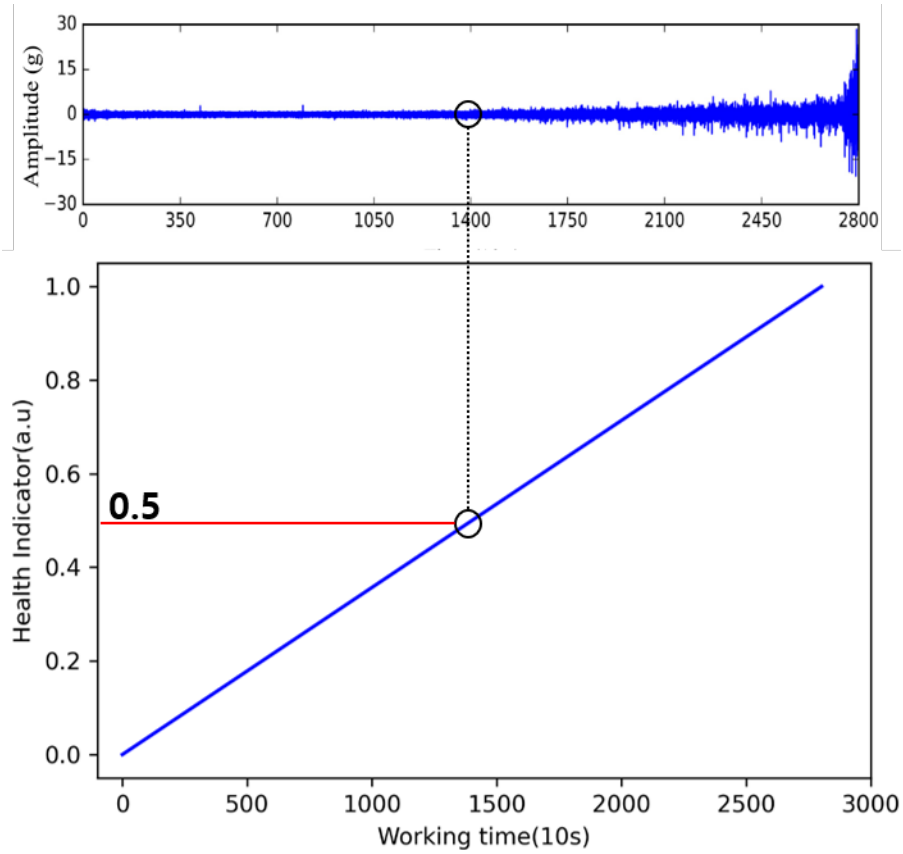

Fig. 7 HI Construction

4 Experiment

4.1 Dataset

The data used in the experiment to validate the methodology is the IEEE PHM challenge PRONOSTIA dataset in 2012 [40]. The PRONOSTIA data consists of 3 working conditions, and the number of bearings in each experimental condition is 7, 7, and 3. In addition, the sampling frequency of each bearing is 25.6kHz, and accelerometers are attached to the x and y axes, and the vibration of the two axes can be obtained. Finally, it consists of a run-to-failure dataset that collected from the start of initial operation until the occurrence of corruption. The summary of data is shown in Table 2.

In addition, 3 data are used for the test data required for methodological verification. Each dataset has a tendency as shown in Fig. 11, and a detailed explanation is given in Table 3. It has individual working condition. Bear-

Table 1 Experiment condition

Experiment condition	number of bearing	notation	Experiment info
1800RPM & 4000N	7	Bearing1_1 bearing 1_7	sampling period interval: 10s
1800RPM & 4000N	7	Bearing1_1 bearing 1_7	sampling frequency: 25.6KHz
1800RPM & 4000N	7	Bearing1_1 bearing 1_7	0.1s snapshot, x,y axis accelerometer

Table 2 Experiment condition

Bearing	Description
Bearing1_1	general failure pattern, monotonic increasing amplitude
Bearing2_1	general failure pattern, sudden increasing amplitude
Bearing3_1	general failure pattern, sudden increasing amplitude, outlier

ing1_1, Bearing2_1, and Bearing3_1, in that order, have more and more complex failure patterns with common failure patterns.

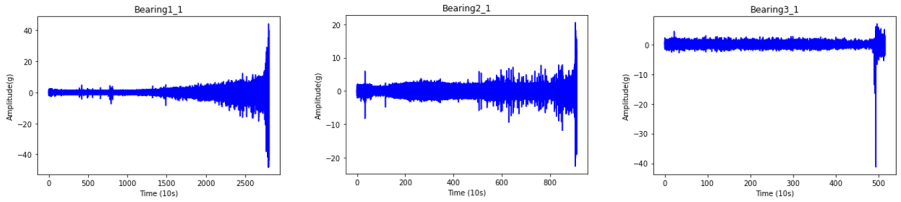

Fig. 8 Model Framework(a) Test Bearing1_1, (b) Test Bearing2_1, (C) Test Bearing3_1

4.2 Model structure

4.3 comparision

4.3.1 model structure

In this section we compare the FEAE model with a single input model. As shown in Figure 12, a comparison is made between using a single input model and using all input data. This is an experiment to evaluate the performance of the model structure.

1. Signal AutoEncoder
2. Statistics AutoEncoder
3. Frequency AutoEncoder

Bearing1_1:
This is the case of the simplest bearing failure pattern with continuous amplitude increase. All models predict HI almost accurately, and Statistics AE predicts the HI closest to the label in a single input model. However, in the

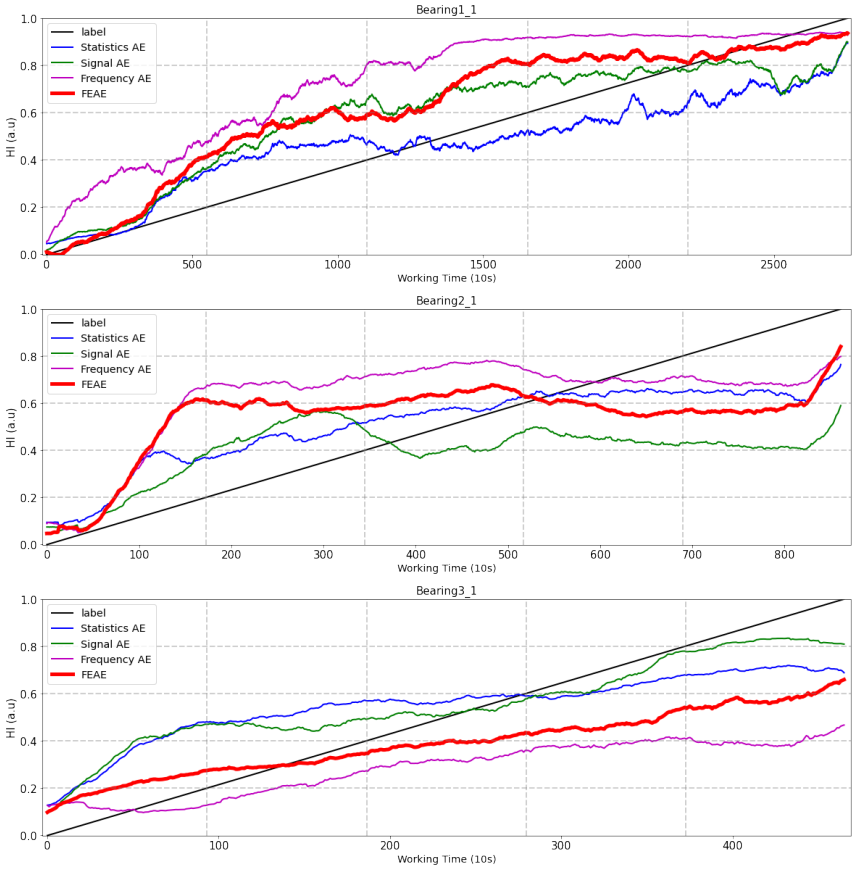

Fig. 9 Experiment of model structure compare to FEAE
(a) Test Bearing1_1, (b) Test Bearing2_1, (c) Test Bearing3_1

case of the Frequency AE model, there is a rapid increase and some differences from other models can be seen. Finally, it can be seen that the Signal AE model looks similar to the proposed FEAE model, but after the second half point, HI decreases again and shows unstable prediction.

Bearing2_1:
Amplitude in the initial state of these, it is a case where it continues until the latter half and has a failure pattern after repeating momentary changes in amplitude. In the initial state, Statistics AE and Signal AE tend to increase in close proximity to label, while Frequency AE and FEAE models tend to increase rapidly. In the middle state, Signal AE tends to decrease suddenly, although the tendency is almost similar. In the latter half, in the case of Statistics AE and Frequency AE, there is a pattern in which HI decreases again in the last part. Due to the characteristics of RUL It is a bad tendency to decrease again. In addition, the proposed FEAE and Signal AE models tend to

increase to the end. In the final RUL part, the FEAE model shows the HI closest to the label and tends to increase to the end.

Bearing3_1:
This data have a pattern in which a value equal to or greater than the occurrence of a momentary change in amplitude occurs after maintaining the amplitude in the initial state. Statistics AE and Signal AE show a downtrend in the last part, showing bad predictions. The Frequency AE and FEAE models tend to be closest to the label, showing a steadily increasing trend.

?? shows the calculated metric that evaluate HI for the 3 models.

$$Mon(X) = \frac{1}{N-1}\left|No.of\ \frac{d}{dx} > 0 - No.of\ \frac{d}{dx} < 0\right| \tag{2}$$

$$Corr(X,T) = \frac{\left|\sum_{i=1}^{N}(X_0^i - \bar{X})(T^i - \bar{T})\right|}{\sqrt{\sum_{i=1}^{N}(X_0^i - \bar{X})\sum_{i=1}^{N}(T^i - \bar{T})}} \tag{3}$$

$$Cri = \frac{Corr + Mon}{2} \tag{4}$$

Feature type	Monotonicity			Trendability			Criteria			Avg. Criteria
	Bearing 1_1	Bearing 2_1	Bearing 3_1	Bearing 1_1	Bearing 2_1	Bearing 3_1	Bearing 1_1	Bearing 2_1	Bearing 3_1	
Signal	0.129	0.206	0.368	0.901	0.549	0.957	0.515	0.377	0.663	0.518
Statistics	0.138	0.203	0.333	0.941	0.906	0.914	0.539	**0.555**	0.624	**0.572**
Frequency	0.102	0.157	0.252	0.900	0.640	0.957	0.501	0.398	0.605	0.501
FEAE	0.147	0.175	0.419	0.932	0.589	0.993	**0.540**	0.382	**0.706**	0.542

Fig. 10 metric

In case of Monotonicity, K is the number of data in the entire life cycle. Monotonicity = 1 mean that there is a complete monotonic, in other cases it means that HI oscillates. There is an irreversible relationship between the actual machine failure tendencies. Failure do not recover spontaneously without human intervention. Appropriate HI Monotonicity tends to increase or decrease monotonically, which is usually the case.

Trendability shows the linear correlation between operating time and HI. At this time, K is the number of data in the entire life cycle, x means HI, and t means the operating time. As the operating time increases, the faulty is gradually failing. Therefore, the trendability of the general situation, that is, the time and HI linear correlation, a value close to 1 will be a meaningful metric.

Criteria means the average value of the both metric. Rather than relying on only one of the two indicators, using the average of both values provides a more reliable indicator.

Summarizing the experiments on the 3 datasets, Statistics AE is performing well on average. And in certain situations, it shows good performance when using a certrain feature set. However, case of the FEAE model, you can see that it maintains the stable tendency in any situation and has strong robustness. These characteristics are the result of efficiently learning the unique information of the vibration signal that only each time, frequency, and signal has.

4.3.2 input structure

The FEAE model uses all feature set combinations, and at the same time, the model is trained through individual AEs that Signal, Statistics, and Frequency each feature sets that have the different meaning. We also use the same dataset to compare with the basic simple concat model in Figure 16. The comparison of both models uses the same data, but you can see a comparison of whether the performance of the models changes depending on how you train them.

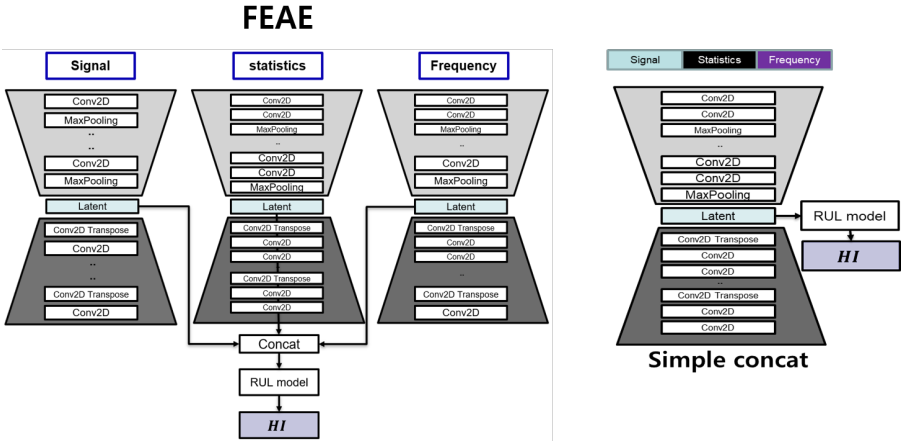

Fig. 11 Network models of 1-stage FEAE(left) and Simple concat(right)

Bearing1_1:
There is no big difference between the FEEA and Simple concat model

Bearing2_1:
Bearing2_1 is The same aspect is shown in the case of Bearing1_1. but there are a few differences. The Simple concat model shows a increase in the initial

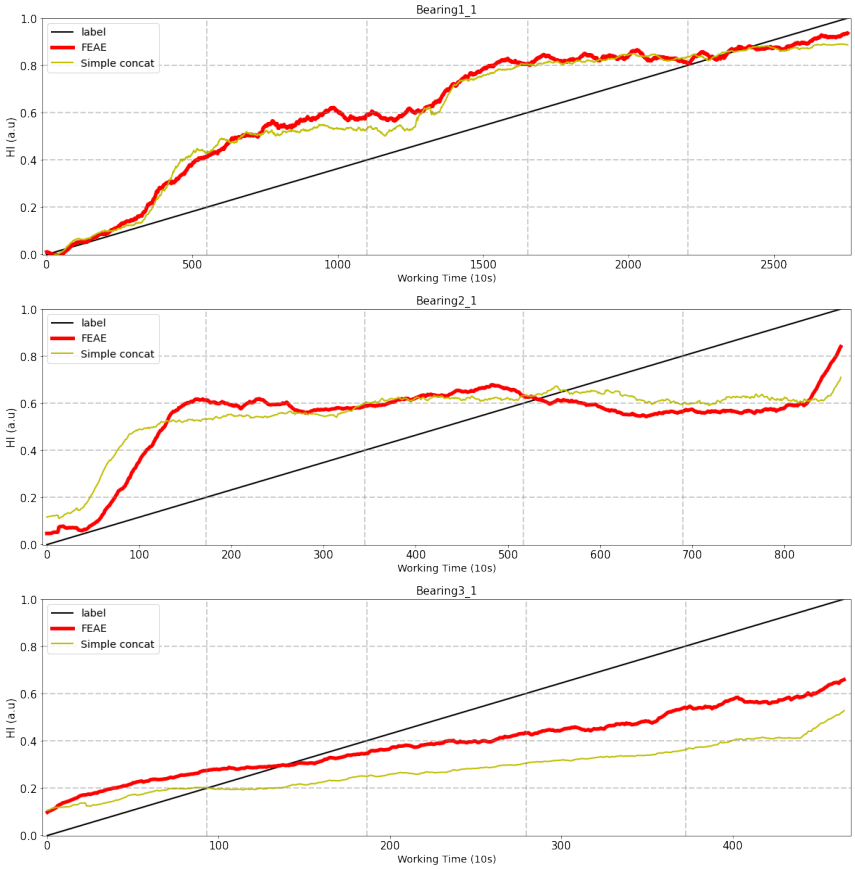

Fig. 12 (a) Average reconstruction error as a function of the number of groups. (b) Average reconstruction errors of ten scenarios.

state, whereas the FEAE model does not. It can also be seen that the FEAE model shows even better predictive power in the final HI predictive value.

Bearing3_1:
There is a big difference in performance from the previous test bearing. Bearing3_1 is data with outliers, and there is a big difference between the two models in terms of performance. Both models also show a steady rise, but the FEAE model can be seen to perform better.

The difference between the two models performance change depending on how the input data is trained. The Simple concat model and the FEAE model learned the same data. The FEAE model train the features of statistics, signals, and frequency domains through different AEs for each domain. In contrast, the Simple concat model concatenates three domains into a single dataset and uses the merged dataset as the training data. The difference between these

learning methods appears in the performance of models, and the method of learning features as one dataset as in existing research is to inefficiently extract the unique information of each domain that each source has.

4.3.3 1stage and 2stage

FEAE model(1stage model) and two model(2stage model) that includes for feature extraction through AE and prediction of HI are compared in this section. The difference between the 1stage model(FEAE) and the 2stage model is that learning is performed at the same time, or features are extracted using the AE that has been trained, and the RUL model is trained. This experiment is a comparative experiment of how the latnet space can be learned efficiently by the two loss functions.

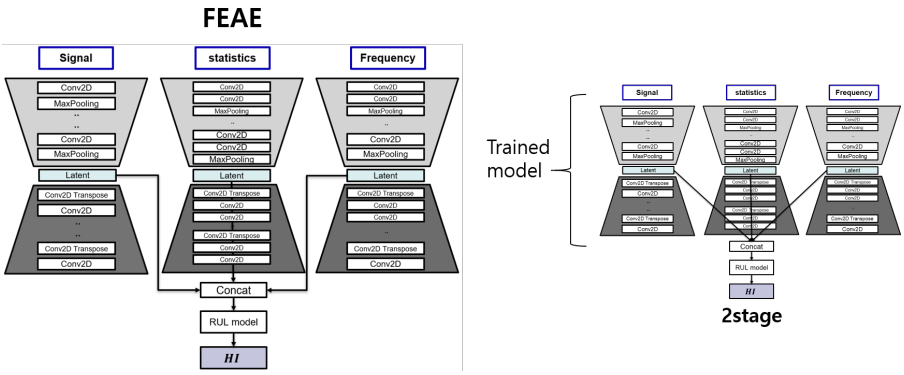

Fig. 13 Network models of 1-stage FEAE(left) and 2-stage FEAE(right)

Bearing1_1:
It looks similar until the initial state, but as time goes on, the 2stage model shows that HI is no longer increasing.

Bearing2_1 & Bearing3_1: Two experiments can see that the initial HI start is non-zero. This can be seen as a result of the vibration signals of the two bearings being generated with the initial amplitude maintained. In addition, the 2stage model does not show a large increase with the passage of time. These features indicate that the latent space of the trained AE does not act as an efficient input feature, and discussions related to training the RUL prediction model apply. be able to.

Learning the RUL prediction model using the latent space of trained AE as an input feature is common to both models (1stage, 2stage). However, the difference between both models is that using a trained AE, it trains the AE as

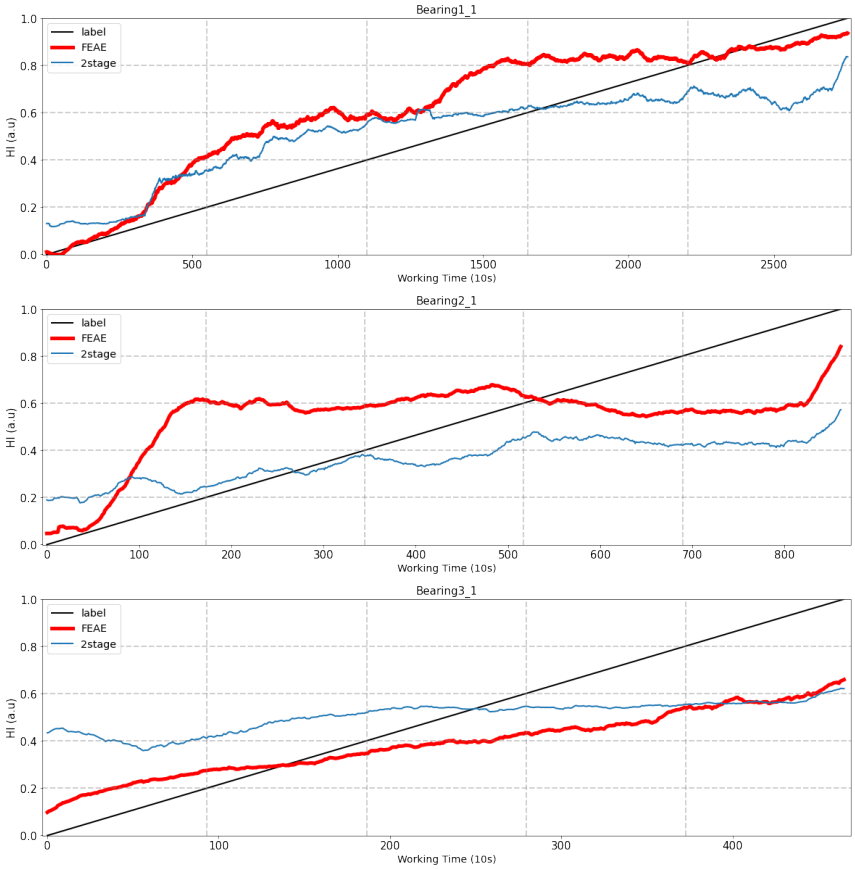

Fig. 14 (a) Average reconstruction error as a function of the number of groups. (b) Average reconstruction errors of ten scenarios.

well as the RUL prediction model. In the case of 1stage model (FEAE), the latent space is trained by two loss function that are the loss function of AE and RUL prediction model. The two loss functions efficiently learn the latent space of each AE through the backpropagation method, which is a learning method of deep learning, and at the same time, the final output of the model has better prediction performance of HI. However, in the case of the 2stage model, the models that predict AE and HI that perform feature extraction have to show poor performance because they are trained separately.

4.3.4 RNN type model

Finally, a comparison was made between the AE-based model and the RNN-based model. RNN-based models have been used to handle time-series data traditionally and are one of the used classic methods. This experiment de-

scribes the difference between a model that uses the latent space of AE and a model that predicts HI using an existing RNN-based model.

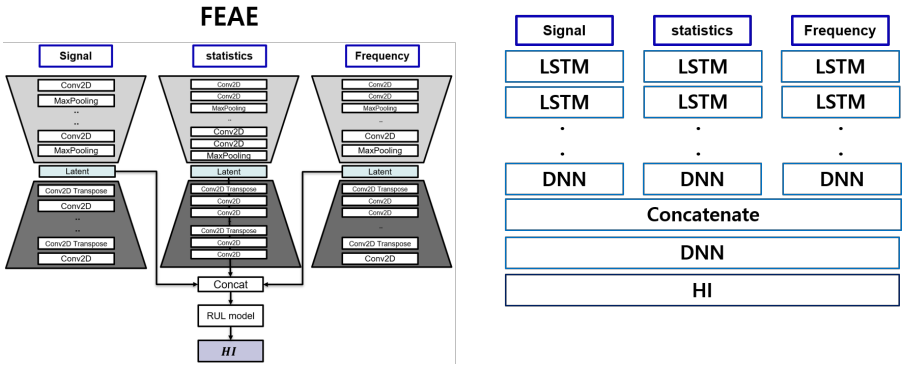

Fig. 15 Network models of 1-stage FEAE(left) and RNN(right)

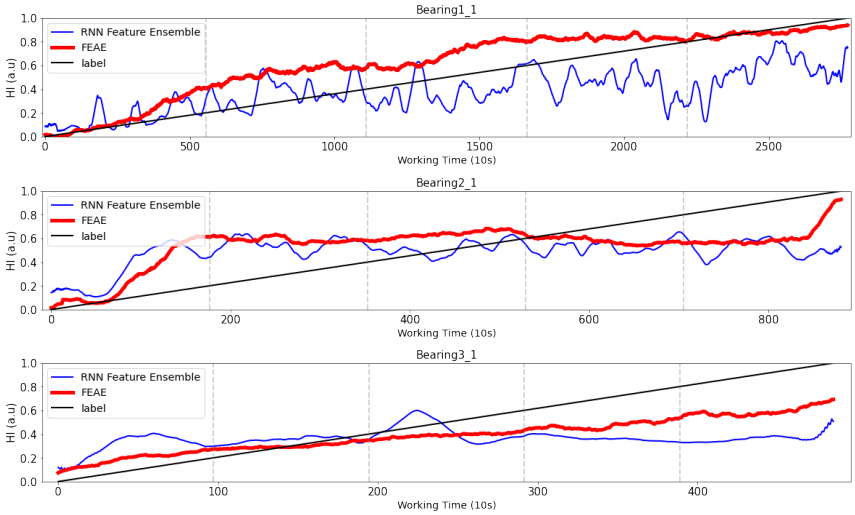

Fig. 16 (a) Average reconstruction error as a function of the number of groups. (b) Average reconstruction errors of ten scenarios.

Bearing1_1:
In the case of the RNN model, there is continuous oscillation compared to the FEAE model. There is a continuous increasing trend until the second half. However, in the end, there is a tendency for it to decrease as it progresses, and then to increase again.

Bearing2_1:
The RNN model starts at 0.2 and can be seen to look similar to FEAE until the second half. However, it can be seen that the FEAE model increases in the part that predicts the final HI, but the RNN model does not increase, but predicts a constant HI.

Bearing3_1:
In the initial state, the RNN model and the FEAE model have the same tendency, but it can be seen that the RNN model does not rise over time as it goes beyond the middle state. And finally the RNN model can be seen to increase a little at the end.

In the case of the FEAE model and the RNN model, the difference is that they are the AE-based model and the RNN model. A model using the latent space of AE can show higher performance than an RNN model using the data of each domain as a feature without performing feature extraction. It can also be seen that the FEAE model increases more stably than the RNN model.

5 Conclusion

In this study, we proposed a methodology for solving the problems related to the selection and extraction of input data and features that should be considered in bearing life prediction. As a method to solve the problem of input data, the signal processing technique classically used for vibration signal analysis was applied. In addition, vibration signals divided each frequency band via WPD, and the each domain was pre-processed to propose method in various studies was used. We applied feature extraction through AE to solve the problem of feature selection. By integrating the 2stage model, which is a drawback of the model using AE, into 1stage, we were able to solve the complicated problem displayed when using the two models, and obtain a model with high performance. Experiments were performed using the PRONOSTIA dataset to verify this methodology. Through the FEAE model presented in this study, it was confirmed that higher performance was obtained than when a single AE was used. And the FEAE model made it possible to learn the meaning of the feature more efficiently than when the functions of each area were created as one data set. In addition, feature extraction through AE was efficient. By converting the 2stage model to the 1stage model, which is a drawback of the model using these structures, we were able to obtain a model with even better performance. There are various facilities and working conditions at the site. However, there are many difficulties in applying one algorithm to each facility. If there is an area that best describes each facility, and if the working conditions change, the above problems can occur. However, when applied to equipment via the proposed method, it will be possible to create models that are even stronger or have better performance through ensembles in various areas. When predicting the remaining life based on these advantages, field

workers can obtain more reliable results. Research for predictive maintenance of bearings is the main field of prediction of remaining life. However, the direction of research to be pursued now will be important to analyze the value of HI calculated by any factor through explainable artificial intelligence. As a first step for that, it will be necessary to analyze the latent space of AE. Then, more information can be obtained from the data by analyzing how each input function is related to the latent space and how the HI value is interpreted.

PHM ..... Prognostics and Health Management
PdM ..... Predictive Maintenance
RUL ..... Remaining Useful Life
HI ..... Health Indicator
WT ..... Wavelet Transformation
DWT ..... Discrete Wavelet Transformation
FT ..... Fourier Transformation
WPD ..... Wavelet Packet Decomposition
AE ..... AutoEncoder
FEAE ..... Feature Ensemble AutoEncoder

6 Acknowledgements

This research was supported by the Ministry of Trade, Industry Energy(MOTIE), Korea Institute for Advancement of Technology(KIAT) through the project of Development of Customized Smart HMI Systems (No.20012807)

References

1. Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., Siegel, D.(2014). Prognostics and health management design for rotary machinery sys-tems—Reviews, methodology and applications. Mechanical systems and signal processing, 42(1-2), 314-334.
2. Vichare, N. M., Pecht, M. G.(2006). Prognostics and health management of electronics. IEEE transactionson components and packaging technologies, 29(1), 222-229.
3. Yang, Y., Yao, D., Liu, X. (2020, July). Remaining Useful Life Prediction Based on Stacked Sparse Autoencoder and Echo State Network. In 2020 39th Chinese Control Conference (CCC) (pp. 5922-5926). IEEE.
4. Kan, M. S., Tan, A. C., Mathew, J. (2015). A review on prognostic techniques for non-stationary and non-linear rotating systems. Mechanical Systems and Signal Processing, 62, 1-20.
5. Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., Siegel, D. (2014). Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. Mechanical systems and signal processing, 42(1-2), 314-334.
6. Zarei, J., Poshtan, J. (2007). Bearing fault detection using wavelet packet transform of induction motor stator current. Tribology International, 40(5), 763-769.
7. Motahari-Nezhad, M., Jafari, S. M. (2021). Bearing remaining useful life prediction under starved lubricating condition using time domain acoustic emission signal processing. Expert Systems with Applications, 168, 114391.
8. Pecht, M., Gu, J. (2009). Physics-of-failure-based prognostics for electronic products. Transactions of the Institute of Measurement and Control, 31(3-4), 309-322.

9. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213-237.
10. Zhang, S., Zhang, S., Wang, B., Habetler, T. G. Machine learning and deep learning algorithms for bearing fault diagnostics—A comprehensive review. arXiv 2019. arXiv preprint arXiv:1901.08247.
11. Ren, L., Sun, Y., Cui, J., Zhang, L. (2018). Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. Journal of Manufacturing Systems, 48, 71-77.
12. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical systems and signal processing, 104, 799-834.
13. Wu, J., Wu, C., Cao, S., Or, S. W., Deng, C., Shao, X. (2018). Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines. IEEE Transactions on Industrial Electronics, 66(1), 529-539.
14. Yang, B., Liu, R., Zio, E. (2019). Remaining useful life prediction based on a double-convolutional neural network architecture. IEEE Transactions on Industrial Electronics, 66(12), 9521-9530.
15. Zhang, B., Zhang, S., Li, W. (2019). Bearing performance degradation assessment using long short-term memory recurrent network. Computers in Industry, 106, 14-29.
16. Akpudo, U., Hur, J. W. (2020). A deep learning approach to prognostics of rolling element bearings. International Journal of Integrated Engineering, 12(3), 178-186.
17. Ahmad, W., Khan, S. A., Kim, J. M. (2017). A hybrid prognostics technique for rolling element bearings using adaptive predictive models. IEEE Transactions on Industrial Electronics, 65(2), 1577-1584.
18. Goyal, D., Choudhary, A., Pabla, B. S., Dhami, S. S. (2019). Support vector machines based non-contact fault diagnosis system for bearings. Journal of Intelligent Manufacturing, 1-15.
19. Essien, A., Giannetti, C. (2020). A deep learning model for smart manufacturing using convolutional LSTM neural network autoencoders. IEEE Transactions on Industrial Informatics, 16(9), 6069-6078.
20. Wang, B., Lei, Y., Li, N., Yan, T. (2019). Deep separable convolutional network for remaining useful life prediction of machinery. Mechanical Systems and Signal Processing, 134, 106330.
21. Xu, F., Yang, F., Fan, X., Huang, Z., Tsui, K. L. (2020). Extracting degradation trends for roller bearings by using a moving-average stacked auto-encoder and a novel exponential function. Measurement, 152, 107371.
22. Xia, M., Li, T., Shu, T., Wan, J., De Silva, C. W., Wang, Z. (2018). A two-stage approach for the remaining useful life prediction of bearings using deep neural networks. IEEE Transactions on Industrial Informatics, 15(6), 3703-3711.
23. Li, X., Zhang, W., Ding, Q. (2019). Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. Reliability Engineering System Safety, 182, 208-218.
24. Zhu, J., Chen, N., Peng, W. (2018). Estimation of bearing remaining useful life based on multiscale convolutional neural network. IEEE Transactions on Industrial Electronics, 66(4), 3208-3216.
25. She, D., Jia, M., Pecht, M. G. (2020). Sparse auto-encoder with regularization method for health indicator construction and remaining useful life prediction of rolling bearing. Measurement Science and Technology, 31(10), 105005.
26. Guo, L., Li, N., Jia, F., Lei, Y., Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing, 240, 98-109.
27. Ren, L., Cheng, X., Wang, X., Cui, J., Zhang, L. (2019). Multi-scale dense gate recurrent unit networks for bearing remaining useful life prediction. Future Generation Computer Systems, 94, 601-609.
28. Sadoughi, M., Lu, H., Hu, C. (2019, June). A Deep Learning Approach for Failure Prognostics of Rolling Element Bearings. In 2019 IEEE International Conference on Prognostics and Health Management (ICPHM) (pp. 1-7). IEEE.

29. Wang, B., Lei, Y., Li, N., Li, N. (2018). A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. IEEE Transactions on Reliability, 69(1), 401-412.
30. Hu, C. H., Pei, H., Si, X. S., Du, D. B., Pang, Z. N., Wang, X. (2019). A prognostic model based on DBN and diffusion process for degrading bearing. IEEE Transactions on Industrial Electronics, 67(10), 8767-8777.
31. Mosallam, A., Medjaher, K., Zerhouni, N. (2016). Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. Journal of Intelligent Manufacturing, 27(5), 1037-1048.
32. Kundu, P., Chopra, S., Lad, B. K. (2019). Multiple failure behaviors identification and remaining useful life prediction of ball bearings. Journal of Intelligent Manufacturing, 30(4), 1795-1807.
33. Bravo-Imaz, I., Ardakani, H. D., Liu, Z., García-Arribas, A., Arnaiz, A., Lee, J. (2017). Motor current signature analysis for gearbox condition monitoring under transient speeds using wavelet analysis and dual-level time synchronous averaging. Mechanical Systems and Signal Processing, 94, 73-84.
34. Lin, P., Tao, J. (2019, June). A novel bearing health indicator construction method based on ensemble stacked autoencoder. In 2019 IEEE international conference on prognostics and health management (ICPHM) (pp. 1-9). IEEE.
35. Shao, H., Jiang, H., Zhao, H., Wang, F. (2017). A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. Mechanical Systems and Signal Processing, 95, 187-204.
36. Safara, F., Doraisamy, S., Azman, A., Jantan, A., Ramaiah, A. R. A. (2013). Multi-level basis selection of wavelet packet decomposition tree for heart sound classification. Computers in biology and medicine, 43(10), 1407-1414.
37. Yan, R., Gao, R. X., Chen, X. (2014). Wavelets for fault diagnosis of rotary machines: A review with applications. Signal processing, 96, 1-15.
38. Altmann, J., Mathew, J. (2001). Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis. Mechanical systems and signal processing, 15(5), 963-977.
39. Dolabdjian, C., Fadili, J., Leyva, E. H. (2002). Classical low-pass filter and real-time wavelet-based denoising technique implemented on a DSP: a comparison study. The European Physical Journal Applied Physics, 20(2), 135-140.
40. Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., Varnier, C. (2012, June). PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In IEEE International Conference on Prognostics and Health Management, PHM'12. (pp. 1-8). IEEE Catalog Number: CPF12PHM-CDR.