

A Deep Learning-Based Remaining Useful Life Prediction Approach for Bearings

Cheng Cheng¹, Guijun Ma¹, Yong Zhang¹, Mingyang Sun¹, *Member, IEEE*,
Fei Teng², *Member, IEEE*, Han Ding, *Senior Member, IEEE*, and Ye Yuan¹, *Member, IEEE*

Abstract—In industrial applications, nearly half the failures of motors are caused by the degradation of rolling element bearings (REBs). Therefore, accurately estimating the remaining useful life (RUL) for REBs is of crucial importance to ensure the reliability and safety of mechanical systems. To tackle this challenge, model-based approaches are limited by the complexity of mathematical modeling. Conventional data-driven approaches, on the other hand, require massive efforts to extract the degradation features and construct the health index. In this article, a novel data-driven framework is proposed to exploit the adoption of deep convolutional neural networks (CNNs) in predicting the RULs of bearings. More concretely, raw vibrations of training bearings are first processed using the Hilbert–Huang transform to construct a novel nonlinear degradation energy indicator which can be used as the training label. The CNN is then employed to identify the hidden pattern between the extracted degradation energy indicator and the raw vibrations of training bearings, which makes it possible to estimate the degradation of the test bearings automatically. Finally, testing bearings' RULs are

predicted through using an ϵ -support vector regression model. The superior performance of the proposed RUL estimation framework, compared with the state-of-the-art approaches, is demonstrated through the experimental results. The generality of the proposed CNN model is also validated by performance test on other bearings undergoing different operating conditions.

Index Terms—Convolutional neural networks (CNNs), Hilbert–Huang transform (HHT), remaining useful life (RUL) estimation, rolling bearings.

NOMENCLATURE

CNN	Convolutional neural network.
DEI	Degradation energy indicator.
EMD	Empirical mode decomposition.
ETA	Exponential transformed accuracy.
HHT	Hilbert–Huang transform.
IMF	Intrinsic mode function.
MAE	Mean average error.
NRMSE	Normalized root-mean-square error.
MSE	Mean-square error.
REB	Rolling element bearing.
RUL	Remaining useful life.
SVR	Support vector regression.
FT or L_{ft}	Failure threshold.
N	Length of historical units of training bearing.
Q	Length of historical units of test bearing.
U	Number of predicted units of test bearing.
S_i	Sensor measurement signal of i th unit.
Δt	Sampling time.
P	Number of measurements in i th unit.
τ	Time interval between two recording phases.
L	DEI of training bearing.
L^{norm}	Normalized DEI of training bearing.
L^{test}	Estimated DEI of test bearing.
$\hat{L}^{\text{U, test}}$	Predicted DEI of test bearing.
K	Total number of layers.
W^k	Weights in k th convolutional layer.
B^k	Bias in k th convolutional layer.
\hat{T}_{failure}	Predicted RUL of test bearing.
T_{failure}	Real RUL of test bearing.
β	Number of test bearings.
$E_r\%$	Relative percentage error of prediction.
S_{mean}	Average score of prediction.
X	Training set for SVR.

Manuscript received August 19, 2019; accepted January 23, 2020. Date of publication February 4, 2020; date of current version June 15, 2020. Recommended by Technical Editor J. Scruggs. This work was supported in part by the National Natural Science Foundation of China under Grant 91748112 and in part by the Primary Research and Development Plan of Jiangsu Province under Grant BE2017002. (Corresponding author: Ye Yuan.)

C. Cheng is with the Key Laboratory of Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: c_cheng@hust.edu.cn).

G. Ma and H. Ding are with the State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China, and also with the School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: mgj@hust.edu.cn; dinghan@hust.edu.cn).

Y. Zhang is with the School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China (e-mail: zhangyong77@wust.edu.cn).

M. Sun is with the College of Control Science and Engineering, Zhejiang University, Hangzhou 310007, China (e-mail: mingyangsun@zju.edu.cn).

F. Teng is with the Department of Electrical and Electronic Engineering, Imperial College London SW7 2AZ, London, U.K. (e-mail: f.teng@imperial.ac.uk).

Y. Yuan is with the Key Laboratory of Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China, and also with the State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: yye@hust.edu.cn).

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

Digital Object Identifier 10.1109/TMECH.2020.2971503

- \mathbb{R} Set of real numbers.
 \mathbb{Z} Set of positive integers.

I. INTRODUCTION

TO CONSTRAIN relative motion while reducing friction between moving parts, rolling element bearings (REBs) are one of the most widely used elements in industrial machinery. Prognostics and health management of bearings is of significance for safety, reliability, and effectiveness of the mechanical systems [1], [2]. In [3], it is shown that nearly half of motor failures are related to the degradation of bearings. As such, estimating the remaining useful life (RUL) (i.e., time-to-failure prognostics) of bearings has attracted a great deal of attention in recent years [4]. RUL prediction helps users monitor the condition of the bearings and provides an estimation of time left before a failure occurs. Compared with fault diagnosis, which has been well investigated over past few decades [5], the problem of RUL prediction studied in this article is a relatively new and challenging topic due to the huge amount of uncertainties of environment and operating condition.

In general, RUL prediction approaches can be categorized into model-based and data-driven approaches. Model-based approaches aim to build a physical model to represent the degradation of the rolling bearing [6]. Li *et al.* [7] predicted the defect growth on a bearing unit using Paris's law for fatigue. However, it is difficult to construct a precise physical degradation model due to the sensitivity of the model parameters and noised operating environments. This limits the practical applications of the model-based approaches. On the other hand, data-driven approaches benefit from the extensive expertise in signal processing and machine learning [8], and infer the degradation process of bearings without knowing any physics of degradation failure. The prognostic framework of the data-driven approaches mainly consists of three stages: 1) feature extraction from noisy sensory signals, which helps to build up the health indicator for the learning of system degradation behavior; 2) degradation models are trained on the training bearing using statistical or machine learning techniques; and 3) the degradation indicator of the test bearings can be estimated based on the model trained in the second stage. Then, the unknown degradation process can be predicted by applying regression techniques (i.e., ϵ -SVR).

To extract features from raw signals, time-domain, frequency-domain, and time-frequency-domain analysis are commonly adopted. Among them, the time-frequency analysis has been found to be the most efficient due to its ability to characterize transient signals over time and frequency domains [9]. Well known time-frequency techniques for extracting bearing features include short-time Fourier transform [10], wavelets [11], Wigner-Ville distribution [12], and Hilbert-Huang transform (HHT) [13]. Implementation of the short-time Fourier transform is limited by its time-frequency resolution capability; for instance, low frequencies are difficult to identify with short windows. On the other hand, wavelets and the Wigner-Ville distribution provide richer pictures than short-time Fourier transform; however, their effectiveness depends on estimating the

Hurst parameter and the quality of the analyzed signal. HHT shows better computational efficiency and resolution over other time-frequency analysis which uses the techniques of empirical mode decomposition (EMD) and Hilbert transform (HT) to decompose the original vibration signal into a number of intrinsic mode functions (IMFs) in various frequency scales. Frequency components of each IMF are related to both the sampling frequency and the signal itself, thus demonstrating that HHT is a self-adaptive signal processing technique perfectly suited to nonstationary signals. Wu *et al.* [14] analyzed the time-to-failure prognostics of REBs, which extracts ten statistical features using time and frequency analysis and 11 IMF features using HHT time-frequency analysis. The gear fault identification method proposed in [15] is based on the HHT, and the first six IMFs are selected as inputs for SOM neural networks for fault diagnosis.

Data-driven RUL prediction approaches are mainly based on statistical and machine learning techniques, such as artificial neural networks (ANN) [16], fuzzy logic systems [17], and autoregressive (AR) models [18]. The computational cost of ANN is relatively high in terms of optimizing the weights of the model. The performances of the AR models and fuzzy logic systems require a precise trend of historical observations and high-quality training data, respectively. Recently, deep learning has merged into research and industry fields, and has beaten other machine learning techniques in speech recognition and image recognition tasks [19]. Deep learning model is good at discovering high-level abstractions from labeled data using a back-propagation algorithm [20]. Specifically, it learns feature representations automatically rather than designing the hand-created features by experience. As the most well known model in deep learning, in recent years, CNN dominates the recognition and detection problems in computer vision domain, which is distinguished by three characteristics, namely, local connections, shared weights, and local pooling [21], [22]. The first two characteristics show that the CNN model requires fewer parameters to detect local information than multilayer perceptron, while the last characteristic ensures shift invariance to the networks. Typically, one-dimensional (1-D) CNN will be employed to this article to learn the latent space of input sensory time-series vibrations, which has been applied with great success to speech recognition and document reading tasks. Few attempts have been made for the prediction problem using CNN-based models [23]–[26]. This article exploits the adoption of CNN technique in estimating the RUL of bearings, as a prognosis problem, to learn about the nonlinear degradation behavior according to raw vibration data and an extracted label. Instead of using the CNN technique to perform the time-series prediction, the main function of the CNN model in this article is to reveal the hidden dependencies between the vibration data and the DEI of the training bearing, which makes full use of the advantages of CNN in automatic feature extraction.

In this article, we propose a data-driven framework for predicting the RUL of REBs by applying the HHT, CNN, and ϵ -SVR. The raw vibration signals collected from sensors are processed by the HHT method and a novel time-series degradation indicator, i.e., DEI, is constructed. Subsequently, a CNN

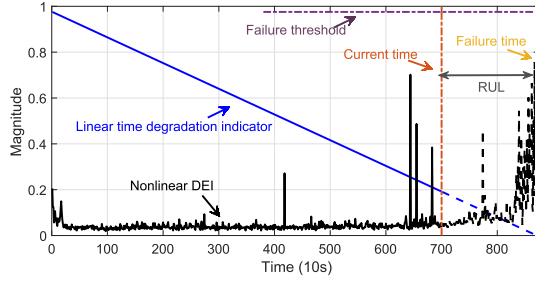


Fig. 1. Linear time degradation indicator (blue) versus nonlinear DEI (black), as a label for network training. The nonlinear DEI experiences a long time flat curve before a sharp degradation trend when it gets close to the end of bearing lifetime, which reflects the real degradation process of most machinery systems [27]. With the aggregation of damages at different bearing components close to the end of lifetime, the simple time degradation indicator is less effective than DEI for RUL estimation.

model is trained to learn the features from the input raw vibration to the DEI label on the training bearings, and used to predict the DEIs of testing bearings. Then, an ϵ -SVR model is introduced so that the evolution of the degradation can be forecast till the bearing failure. The effectiveness of the proposed framework for RUL prediction is validated on an experimental platform (i.e., PRONOSTIA). Much lower RUL prediction errors are achieved, compared with eight existing approaches in previous papers and two tested methods designed in this article, indicating the superior performance of the proposed method.

This article makes the following contributions.

- 1) The proposed method successfully extracts a novel nonlinear DEI (see Fig. 1, compared with the linear time degradation indicator) to describe the degradation trend of the training bearing, according to the natural frequencies of bearing components.
- 2) The proposed CNN architecture is general and robust for similar operating conditions—it can transfer to another bearing undergoing different operating conditions and obtain good prediction results, without changing CNN hyperparameters and the depth of layers.
- 3) The proposed DEI is an integrated indicator with regard to the maximum vibration levels among different bearing components which considers all the possible defects on the REB. This is a more realistic indicator as the localized defects are not initially initiated in real industrial applications, meaning that all the types of defects have to be considered.
- 4) CNN scales all the indicators of training bearings and test bearings into a consistent latent space. Thus, training and testing can share the same failure threshold (FT), i.e., the maximum value of the indicator for the training bearing.

The rest of this article is organized as follows. Section II presents the proposed RUL prediction framework with technical details. In Section III, experimental results obtained from

bearing degradation tests are carried out. Thereby, the performance of the proposed framework is validated and the results show improved accuracy in predicting the RUL compared with eight state-of-the-art approaches and two designed test methods in this article. Finally, Section IV concludes this article.

II. DEGRADATION INDICATOR TRAINING AND RUL PREDICTION ALGORITHM

The overall framework for the prediction of the RUL can be decomposed into three parts. The schematic of the overall framework is shown in Fig. 2. The key challenges of this article involve: 1) obtain the DEI to represent the degradation behaviour; 2) establish a CNN model to map raw vibration signal to the DEI; and 3) construct an ϵ -SVR to predict the RUL. Thus, in the following subsections, the explicit expression of degradation feature extraction, CNN model, and ϵ -SVR forecasting model will be derived in Sections II-A–II-C.

A. Degradation Indicator Extraction

To begin with, for a training bearing, it is assumed that the raw vibration signal till the end of lifetime with $N \in \mathbb{Z}$ historical units have been acquired. The sensory vibration signal $\mathbf{S}_i = \{S_i(t_p)\}_{p=1}^P$, with sampling time Δt , is measured at each historical unit for $i \in D = \{1, 2, \dots, N\}$, where P is the number of measurements recorded in each historical unit.

EMD is a self-adaptive method which is normally applied to analyze nonstationary and nonlinear signals. It decomposes the raw vibration data \mathbf{S}_i into n number of IMFs, illustrating the natural oscillation modes from fast to low oscillations.

For the i th unit, the j th mode ($j = 1, \dots, n$) of IMF, $\text{IMF}_{i,j}(t_p) = a_{i,j}^v(t_p)$ for $p = 1, \dots, P$, is calculated iteratively associated with the iteration number $v = 0, 1, \dots$

First, let $v = 0$ and initialize $a_{i,j}^0(t_p)$ for $p = 1, \dots, P$ by

$$a_{i,j}^0(t_p) = \begin{cases} S_i(t_p), & j = 1 \\ S_i(t_p) - \sum_{j=1}^{n-1} \text{IMF}_{i,j}(t_p), & j = 2, \dots, n \end{cases} \quad (1)$$

Define $\text{IMF}_{i,j}(t_p) = a_{i,j}^0(t_p)$ for $p = 1, \dots, P$ only if the IMF meets the following two conditions.

- I) The IMF should have one or zero difference between the extrema number and the number of zero crossing.
- II) Along the time axis, the average value of upper and lower bound of the IMF should be zero everywhere.

Otherwise, update the $\text{IMF}_{i,j}(t_p) = a_{i,j}^v(t_p)$ for $p = 1, \dots, P$ through the following iteration procedure with $v = 1, 2, \dots$ until the IMF satisfies both the aforementioned conditions (I) and (II)

$$a_{i,j}^v(t_p) = a_{i,j}^{v-1}(t_p) - \omega_{i,j}^{v-1}(t_p), \quad \forall p \quad (2)$$

where $\omega_{i,j}^{v-1}(t_1), \dots, \omega_{i,j}^{v-1}(t_P)$ are the mean values of the upper envelope and the lower envelope of $a_{i,j}^{v-1}(t_1), \dots, a_{i,j}^{v-1}(t_P)$.

Once obtaining all IMFs, the analytical form of each IMF can be written as

$$\text{IMF}_{i,j}^A(t_p) = \text{IMF}_{i,j}(t_p) + i \text{IMF}_{i,j}^H(t_p) \quad (3)$$

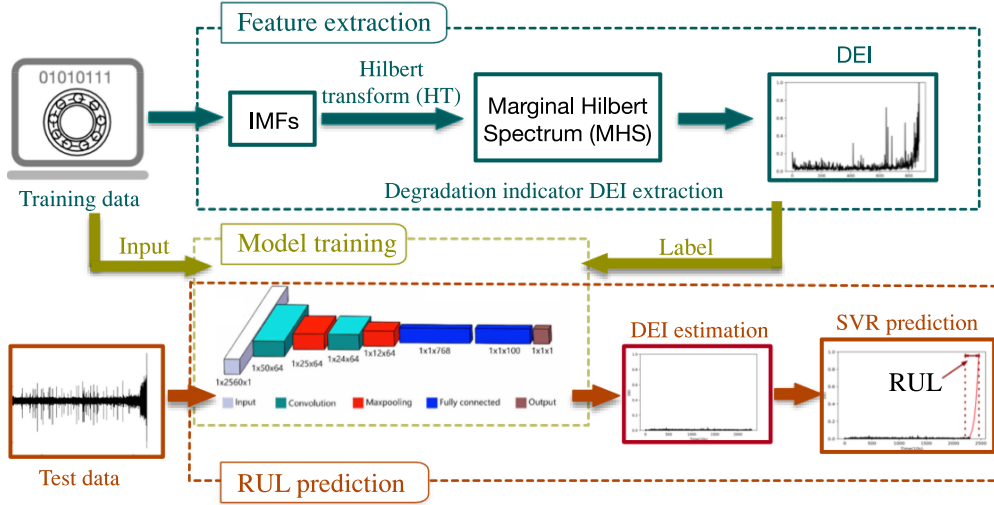


Fig. 2. Data-driven RUL prediction framework. This framework involves three stages: 1) a feature extraction stage aims to extract the degradation indicator. The DEI is extracted using EMD, HT, and MHS. The value of DEI is also related to the natural frequencies of different bearing components; 2) a model training stage aims to train a CNN model based on the training data that discovers the hidden pattern between the DEI and the raw vibration data; and 3) a prediction stage predicts the RUL according to the trained CNN model and an ϵ -SVR forecasting model.

where \mathbf{i} is the imaginary part of the $\text{IMF}_{i,j}^A(t_p)$. $\text{IMF}_{i,j}^H(t_p)$ is the HT of $\text{IMF}_{i,j}(t_p)$ by convolution with function $\frac{1}{\pi t}$, given as

$$\text{IMF}_{i,j}^H(t_p) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\text{IMF}_{i,j}(s)}{t_p - s} ds. \quad (4)$$

By this means, we can calculate the instantaneous amplitude $h_{i,j}(t_p)$ and phase $\phi_{i,j}(t_p)$

$$h_{i,j}(t_p) = \sqrt{(\text{IMF}_{i,j}(t_p))^2 + (\text{IMF}_{i,j}^H(t_p))^2}$$

$$\phi_{i,j}(t_p) = \arctan\left(\frac{\text{IMF}_{i,j}^H(t_p)}{\text{IMF}_{i,j}(t_p)}\right). \quad (5)$$

Accordingly, it is easy to derive the instantaneous frequency

$$f_{i,j}(t_p) = \frac{1}{2\pi} \frac{\phi_{i,j}(t_p + \Delta t) - \phi_{i,j}(t_p - \Delta t)}{2\Delta t} \quad (6)$$

for $p = 1, \dots, P$.

Then, the Hilbert spectrum of \mathbf{S}_i is obtained by

$$M_i(f_i, t_p) = \sum_{j=1}^n h_{i,j}(f_{i,j}, t_p). \quad (7)$$

The marginal Hilbert spectrum (MHS) $M_i(f_i)$ can be written as

$$M_i(f_i) = \sum_{p=1}^P M_i(f_i, t_p). \quad (8)$$

Natural frequencies of bearing components depend on the geometry of the bearing and its rotation speed. Expression of these frequencies are given in Table I, where η is the number of balls, f_ω is the rotation frequency, Φ is the contact angle, and ℓ_{ball} and ℓ_{pitch} are the ball diameter and pitch diameter, respectively.

With the bearing frequencies of different components, the value of the DEI L_i at historical unit i is defined as the maximum

TABLE I
BEARING FREQUENCIES OF INNER RACE, OUTER RACE, AND BALL [28]

Symbol	Description	Expression
f_{inner}	Inner race frequency	$\frac{\eta}{2} \cdot f_\omega \cdot \left(1 + \frac{\ell_{\text{ball}}}{\ell_{\text{pitch}}} \cdot \cos \Phi\right)$
f_{outer}	Outer race frequency	$\frac{\eta}{2} \cdot f_\omega \cdot \left(1 - \frac{\ell_{\text{ball}}}{\ell_{\text{pitch}}} \cdot \cos \Phi\right)$
f_{ball}	Ball frequency	$\frac{\ell_{\text{ball}}}{\ell_{\text{pitch}}} \cdot f_\omega \cdot \left(1 - \frac{\ell_{\text{ball}}^2}{\ell_{\text{pitch}}^2} \cdot \cos^2 \Phi\right)$

value of the MHS by substituting f_{inner} , f_{outer} , and f_{ball} into $M_i(f_i)$, given that

$$L_i = \max_{f_i \in \{f_{\text{inner}}, f_{\text{outer}}, f_{\text{ball}}\}} M_i(f_i), \quad i \in D. \quad (9)$$

The extracted DEI $\mathbf{L} = [L_1, \dots, L_N]$ is normalized before training

$$L_{\text{norm},i} = \frac{L_i - \min(\mathbf{L})}{\max(\mathbf{L}) - \min(\mathbf{L})} \pm \epsilon, \epsilon \rightarrow 0 \quad (10)$$

for $i = 1, \dots, N$, and ϵ is an infinitesimal that is used to avoid the value of label equal to zero or one. Thus, the normalized DEI is

$$\mathbf{L}_{\text{norm}} = [L_{\text{norm},1}, \dots, L_{\text{norm},N}]. \quad (11)$$

B. DEI Pattern Learning

In this article, layers with repeated components are stacked in a CNN architecture, including convolutional layers, pooling layers, fully connected layers, and a regression layer [21].

Convolutional layer contains organized patches in convolutional layers, each patch is calculated by composing the features of the previous layer through a filter bank with the following equation:

$$u_m^k = W^k * u_e^{k-1} + B^k \quad (12)$$

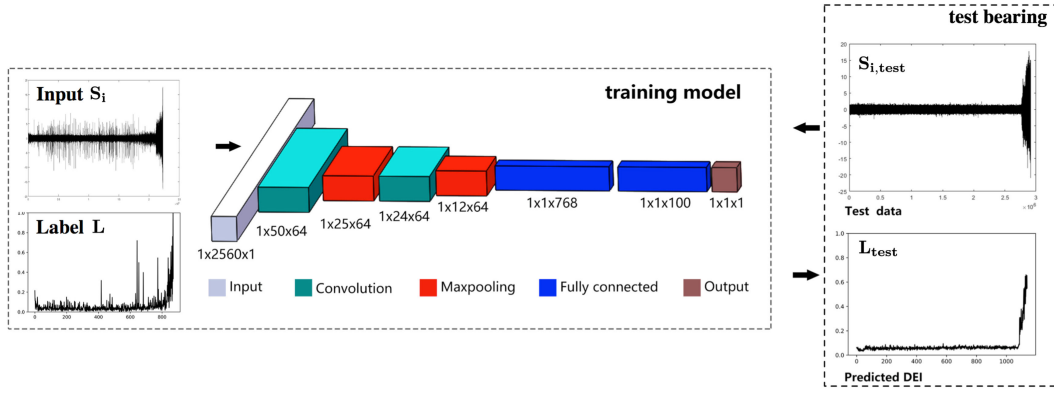


Fig. 3. Diagram of extracting features for ϵ -SVR prediction, where l is the sampling window size, s is the sliding size, and x_g stores the mean value μ_g and variance σ_g^2 of each sampling window.

where $u_m^k \in \mathbb{R}$ denotes the output of the m th unit in the k th layer. $u_e^{(k-1)} \in \mathbb{R}^{1 \times o^k}$ is the input data of the e th subvector in the previous layer $k-1$, where o^k is kernel size in layer k . $W^k \in \mathbb{R}^{1 \times o^k}$ and $B^k \in \mathbb{R}$ denote the connecting weights and bias in the k th layer, respectively. “ $*$ ” means the convolution operation. It is noted that when $k=1$, $u_e^{(k-1)}$ is a subvector of the raw vibration data S_i . Here, we define all neurons in each layer as $\mathbf{u}^k = [u_1^k, \dots, u_m^k, \dots, u_{G^k}^k]$ for $k \in D_k = \{1, 2, \dots, K\}$, where $G^k \in \mathbb{Z}$ is the number of neurons in the k th layer and K is the total number of layers. For convolutional layer, $G^k = (G^{k-1} - o^k)/I_{cv}^k + 1$ and $I_{cv}^k \in \mathbb{Z}$ is the stride in convolutional layer.

Activation function is introduced after convolutional layer. Among various activation functions, rectified linear unit (ReLU) $r_m^k = \max(0, u_m^k)$ is chosen as the nonlinear activation function to prevent the issue of vanishing gradient which may significantly increase the training time or even lead to nonconvergence.

Pooling layer is then used as a nonlinear down-sampling layer to extract the maximum feature values in each patch of the input data. Its function is to save computation time and downsize the number of parameters of the model as well as control overfitting. More specifically, pooling transforms small windows into single values by maxing or averaging. Consequently, the features extracted within the small window are similar and therefore, illustrate the shift invariance property of CNN. Max-pooling layer is selected in this article as it is an algorithmic choice to ensure the generalization of neural networks [29], which is given by

$$P_m^k = \max_{\gamma=1, \dots, \lambda^k} r_{\gamma+(m-1)I_{pl}^k}^{(k-1)} \quad (13)$$

where $\lambda^k \in \mathbb{Z}$ is the pooling size and I_{pl}^k is the stride in max pooling layer.

Fully connected layer and regression layer, like a classic ANN network, take the results of the convolution and max-pooling processes and use them to generate a predicted label. Since we use a normalized DEI $\mathbf{L}_{\text{norm}} \in \mathbb{R}^{1 \times N}$ as the label for learning, the sigmoid function $\text{sigm}(\mathbf{u}^{K-1})$ with an output value between (0,1) is applied to the last layer for normalized output. Hereby, mean-square error (MSE) function is used to compute the loss

Algorithm 1: Outline of CNN Training for DEI Estimation.

Input : The normalized extracted DEI label \mathbf{L}_{norm} ;

The raw vibration data $\{S_i\}_{i=1}^N$.

Output: Trained CNN parameters: W^k and B^k

Initialize parameters;

repeat

Forward Propagation:

do

Conducting convolution operation with the raw vibration data using Eq. (12).

Use ReLU as the nonlinear activation function.

Max-pooling function Eq. (13) is employed to extract the maximum feature values.

end ;

Conventional fully-connected layer is used for DEI regression.

The sigmoid function is introduced for normalized output. Compute the MSE with the loss function using Eq. (14).

Backward Propagation:

Compute the gradient using *Adam* optimizer [30] and update network parameters W^k and B^k .

until *Maximum iterations* ;

Use the trained CNN to estimate the DEI \mathbf{L}_{test} on the test bearing.

with the expression

$$z = \frac{1}{N} \sum_{i=1}^N (L_{\text{norm},i} - \tilde{L}_{\text{norm},i})^2 \quad (14)$$

where the proposed CNN model is minimizing the loss function z between ground label DEI \mathbf{L}_{norm} and predicted label $\tilde{\mathbf{L}}_{\text{norm}}$. Algorithm 1 outlines the proposed CNN modeling procedure.

C. RUL Prediction

With the obtained CNN model, for a new test bearing with $Q \in \mathbb{Z}$ historical units, the estimated DEI $\mathbf{L}_{\text{test}} = [L_{1,\text{test}}, \dots, L_{Q,\text{test}}] \in \mathbb{R}^{1 \times Q}$ can be automatically generated by the trained CNN with the new vibration signal S_i , where $i \in D_{\text{test}} = \{1, \dots, Q\}$. Then, to predict the RUL $\hat{T}_{\text{failure}} \in \mathbb{R}^+$, an ϵ -SVR forecasting model [31] is formalized to predict the

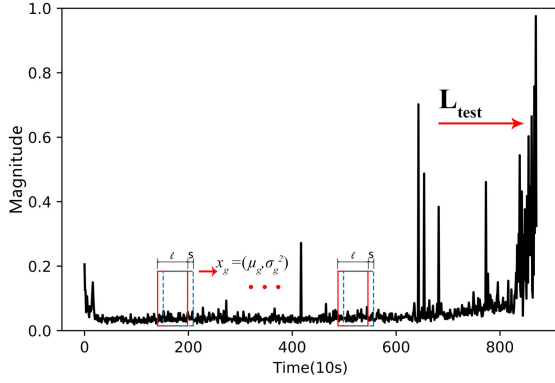


Fig. 4. Overview of the PRONOSTIA platform [32].

upcoming degradation $\hat{L}_{Q+1,\text{test}}, \hat{L}_{Q+2,\text{test}}, \dots$ based on the estimated DEI \mathbf{L}_{test} by sliding window method. The forecasting process contains the following three substeps.

- 1) Extract training features from the estimated DEI \mathbf{L}_{test} over a sliding window. The schematic of this step is illustrated in Fig. 3. The estimated DEI \mathbf{L}_{test} is decomposed into overlapping windows associated with sampling window size l and sliding size s . $x_g = (\mu_g, \sigma_g^2)$ for $g \in \{1, \dots, \frac{Q-l-1}{s} + 1\}$ represents a training feature for ϵ -SVR, where μ denotes the mean value and σ^2 denotes the variance of each sampling window. Thus, the training set for ϵ -SVR $\mathbf{X} = [(x_1, L_{l+1,\text{test}}), \dots, (x_g, L_{(g-1)s+l+1,\text{test}}), \dots, (x_{\frac{Q-l-1}{s}+1}, L_{Q,\text{test}})]$ is obtained, where $L_{(g-1)s+l+1,\text{test}}$ corresponds to the next value in \mathbf{L}_{test} of the g th sampling window.
- 2) ϵ -SVR modeling is described in Algorithm 2, while at the application level, two parameters (distance limit $\epsilon \in \mathbb{R}$ and penalty parameter $C \in \mathbb{R}$) can be set manually when training the prediction model. A radial basis function (RBF) is necessary when we intend to train a nonlinear model.
- 3) The SVR model $f(x)$ learned in Algorithm 2 is then used to predict the RUL by sliding window method (with the same l and s in step 1). Since the test bearing is undergoing the same operating condition as the training bearing, it is reasonable to define the FTs (denoted as L_{ft}) of the test bearing equal to the last feature of the DEI of the training bearing, such that $L_{ft} = L_N$. Hence, the first prediction can be calculated as $\hat{L}_{Q+1,\text{test}} = f(x_{\frac{Q-l-1}{s}+1})$, and the predicted DEI $\hat{\mathbf{L}}_{\mathbf{U},\text{test}} = [\hat{L}_{Q+1,\text{test}}, \dots, \hat{L}_{Q+U,\text{test}}]$ can be obtained by shifting the sampling window, with $\hat{L}_{Q+U-1,\text{test}} \leq L_{ft} \leq \hat{L}_{Q+U,\text{test}}$. This will consequently lead to $\hat{T}_{\text{failure}} = U \times \tau$, where τ is the time interval between two recording phases.

III. EXPERIMENTS

A. Data Description

The validation of the proposed RUL prediction framework is conducted on an experimentation platform named PRONOSTIA (see Fig. 4). This platform is built as a combination of three

Algorithm 2: Framework of ϵ -SVR.

Require:

A training set \mathbf{X} ;
 A distance limit ϵ and a penalty parameter C ;
 A kernel function named RBF with the equation:
 $\kappa(x_g, x_q) = \exp\left(-\frac{\|x_g - x_q\|^2}{2\sigma^2}\right)$, σ is the width of RBF and $\sigma > 0$;

Ensure:

A regression model like $f(x) = w^T \phi(x) + b$, where w and b are optimized parameters and $\phi(x)$ is the x -mapped eigenvector that satisfies the equation:

$$\kappa(x_g, x_q) = \langle \phi(x_g), \phi(x_q) \rangle.$$

Step 1: Establish optimization problems using C , ϵ , \mathbf{X} , and two slack variables ξ_g and $\hat{\xi}_g$ (slack degree for upper boundary and lower boundary, respectively) as following:

$$\begin{aligned} \min_{w, b, \xi_g, \hat{\xi}_g} \quad & \frac{1}{2} \|w\|^2 + C \sum_{g=1}^{\frac{Q-l-1}{s}+1} (\xi_g + \hat{\xi}_g) \quad (i) \\ \text{s.t.} \quad & f(\phi(x_g)) - L_{(g-1)s+l+1,\text{test}} \leq \epsilon + \xi_g \\ & L_{(g-1)s+l+1,\text{test}} - f(\phi(x_g)) \leq \epsilon + \hat{\xi}_g \\ & \xi_g \geq 0, \hat{\xi}_g \geq 0. \end{aligned}$$

Step 2: Add Lagrangian multipliers for each constraint: $\mu_g \geq 0, \hat{\mu}_g \geq 0, \alpha_g \geq 0, \hat{\alpha}_g \geq 0$ and get the Lagrange function of formula (i):

$$\begin{aligned} L(w, b, \alpha, \hat{\alpha}, \xi, \hat{\xi}, \mu, \hat{\mu}) = & \frac{1}{2} \|w\|^2 + C \sum_{g=1}^{\frac{Q-l-1}{s}+1} (\xi_g + \hat{\xi}_g) \\ & - \sum_{g=1}^{\frac{Q-l-1}{s}+1} \mu_g \xi_g - \sum_{g=1}^{\frac{Q-l-1}{s}+1} \hat{\mu}_g \hat{\xi}_g + \sum_{g=1}^{\frac{Q-l-1}{s}+1} \alpha_g (f(\phi(x_g)) \\ & - L_{(g-1)s+l+1,\text{test}} - \epsilon - \xi_g) + \sum_{g=1}^{\frac{Q-l-1}{s}+1} \hat{\alpha}_g (L_{(g-1)s+l+1,\text{test}} \\ & - f(\phi(x_g)) - \epsilon - \hat{\xi}_g). \quad (ii) \end{aligned}$$

Step 3: Substitute $f(\mathbf{X})$ into the formula (ii), we firstly compute the gradient of $L(w, b, \alpha, \hat{\alpha}, \xi, \hat{\xi}, \mu, \hat{\mu})$ with respect to w, b, ξ_g and $\hat{\xi}_g$ and make it equal to zero.

Then get the following results:

$$\begin{aligned} w = & \sum_{g=1}^{\frac{Q-l-1}{s}+1} (\hat{\alpha}_g - \alpha_g) \phi(x_g), \quad 0 = \sum_{g=1}^{\frac{Q-l-1}{s}+1} (\hat{\alpha}_g - \alpha_g), \\ C = & \alpha_g + \mu_g, \quad C = \hat{\alpha}_g + \hat{\mu}_g. \quad (iii) \end{aligned}$$

Step 4: Substitute formula (iii) into formula (ii), and get a dual problem:

$$\begin{aligned} \max_{\alpha, \hat{\alpha}} \quad & \sum_{i=g}^{\frac{Q-l-1}{s}+1} L_{(g-1)s+l+1,\text{test}} (\hat{\alpha}_g - \alpha_g) - \epsilon (\hat{\alpha}_g + \alpha_g) \\ & - \frac{1}{2} \sum_{g,q=1}^{\frac{Q-l-1}{s}+1} (\hat{\alpha}_g - \alpha_g) (\hat{\alpha}_q - \alpha_q) \phi(x_g)^\top \phi(x_q) \\ \text{s.t.} \quad & \sum_{g=1}^{\frac{Q-l-1}{s}+1} (\hat{\alpha}_g - \alpha_g) = 0 \\ & 0 \leq \alpha_g, \hat{\alpha}_g \leq C. \end{aligned}$$

TABLE II
PHYSICAL CHARACTERISTICS AND OPERATION CONDITION

Physical parameter	Value
Number of balls of the bearing (η)	13
Ball diameter of the bearing (ℓ_{ball})	3.5 mm
Pitch diameter of the bearing (ℓ_{pitch})	25.6 mm
Contact angle of the bearing (Φ)	0°
Rotation frequency (f_{ω}), bearing1	1800 r/min
Rotation frequency (f_{ω}), bearing2	1600 r/min
Maximum dynamic load (F), bearing1	4000 N
Maximum dynamic load (F), bearing2	4200 N

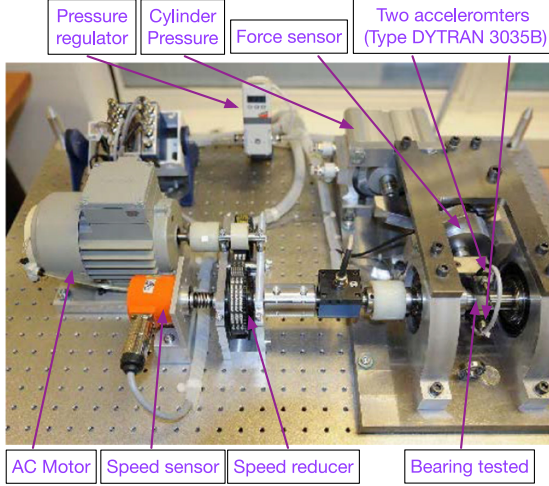


Fig. 5. Evolutions of (a) original DEI and (b) normalized DEI.

Step 5: The above process satisfies the Karush–Kuhn–Tucker (KKT) condition, which means at least one of α_g and $\hat{\alpha}_g$ is equal to 0. A sequential minimization algorithm is used to solve α_g or $\hat{\alpha}_g$. Parameter b is calculated by substituting the α_g to get the mean value. The final regression model is described as:

$$f(x) = \sum_{g=1}^{Q-1} \frac{1}{s} (\hat{\alpha}_g - \alpha_g) \kappa(x, x_g) + b.$$

parts—a rotating part, a loading part, and a measurement part. The rotation of the test bearing is driven by the low-speed shaft whose rotating torque is transmitted from an ac motor. A radial force generated by this loading part is applied on the external ring of the testing bearing. Since this external radial force exceeds the bearing’s allowable dynamic load (4000 N), the degradation behavior is accelerated so that we can observe its degradation process within a relatively short time in few hours. During experimental tests, two high-frequency accelerometers (Type DYTRAN 3035B) are placed orthogonally on the external race of the test bearing to acquire the horizontal and vertical vibrations, respectively. The accelerometer bandwidth is 0.5–10 kHz ($\pm 5\%$), and its natural/resonant frequency is 45 kHz. In this article, we extract our degradation labels by using the horizontal vibrations.

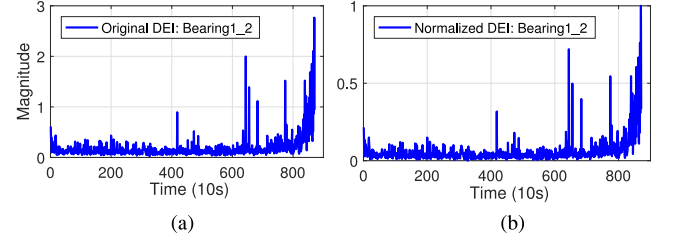


Fig. 6. Time evolutions of the intermedia feature $M_i(f_{inner})$, $M_i(f_{outer})$, and $M_i(f_{ball})$ in (9).

Bearing1 is chosen to validate the proposed framework. More specifically, the training set bearing1_2 is used for extracting the DEI and training the CNN model. Test sets bearing1_4, bearing1_5, and bearing1_6 are then used for estimating their DEIs and predicting the RULs by applying ϵ -SVR forecasting model. Results of Bearing2 under different rotational frequency and external dynamic load are also provided and compared. The geometry parameters and the operation conditions of bearing1 and bearing2 are listed in Table II. Sampling frequency of the vibration sensor is 25.6 kHz. 0.1-s accelerometer vibration signals are recorded at a fixed time interval $\tau = 10$ s. Therefore, each recording phase contains $p = 2560$ measurements. More detailed description of the dataset, bearings, and sensors is presented in the data description in [32].

B. Degradation Indicator Extraction

The DEI of bearing1_2 is extracted by substituting its outer ring frequency $f_{outer} = 168$ Hz, inner ring frequency $f_{inner} = 221$ Hz, and ball frequency $f_{ball} = 215.4$ Hz into (9). The evolution of the final extracted DEI L of bearing1_2 is shown in Fig. 5(a). We also show the time evolutions of intermediate features $M_i(f_{inner})$, $M_i(f_{outer})$, and $M_i(f_{ball})$ of (9) in Fig. 6. It can be observed that the magnitude of each time point in Fig. 5(a) is the maximum of $M_i(f_{inner})$, $M_i(f_{outer})$, and $M_i(f_{ball})$ at that time.

C. Degradation Indicator Estimation

We use the vibration signal in the horizontal direction of the bearing1_2 as the input of the CNN, and the normalized DEI is used as the label which contains historical units $N = 871$. The normalized DEI, L_{norm} , is shown in Fig. 5(b). A less complex architecture of the CNN model is designed to improve the robustness of the network. As shown in Fig. 7, our finalized CNN model consists of $K = 6$ layers—two convolutional layers (Conv1 and Conv2), two max-pooling layers (Maxpooling1 and Maxpooling2), one fully connected layer (FC1), and one regression layer for output. Before model training, *Adam* is set to be the optimizer with a small value 0.00001 as it guarantees a quick loss convergence compared with a larger or smaller learning rate for the CNN training process. The activation function in the output layer is the sigmoid function, while the ReLU function is used in the previous layers. Details of parameters in the proposed CNN model are concluded in Table III. The convolutional window sizes (kernel sizes) of convolutional layers are set to a large

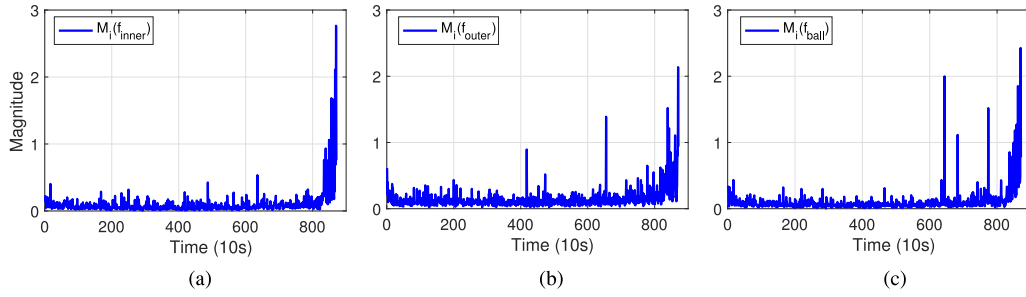


Fig. 7. Proposed CNN architecture. Normalized DEI L is used as the label for training. A six-layer CNN model is trained to map the raw vibration data S_i (input) to the DEI L (output). For a new test bearing, the vibration data $S_{i,test}$ is directly input to the CNN model to obtain the estimated DEI L_{test} .

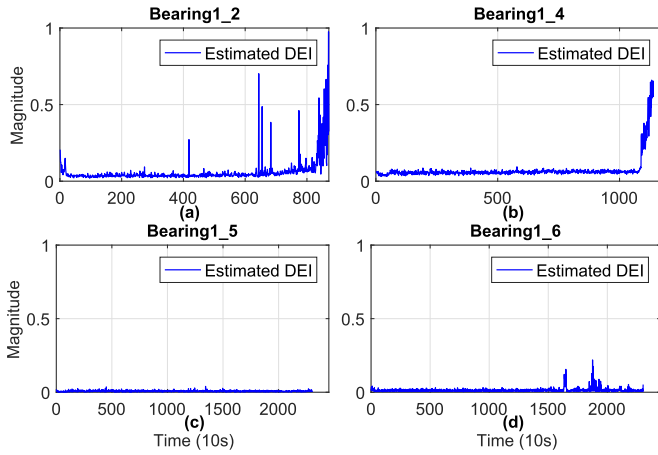


Fig. 8. Estimated DEIs as the output of the proposed CNN model: (a) bearing1_2; (b) bearing1_4; (c) bearing1_5; and (d) bearing1_6.

value 100 and a small value 2, respectively. The kernel size of Conv1 is relevantly large in order to extract more features from the raw vibration signal for more impressive power, meanwhile, small kernel size is selected for Conv2 to prevent overfitting. Hyperparameters are obtained after 1000 iterations of training.

The estimated DEIs L_{test} as the output of the CNN model are shown in Fig. 8. In Fig. 8(a), estimated DEI of the training bearing1_2 shows similar time evolution as the DEI label in Fig. 5(b). The final estimated DEI value of bearing1_2, $L_{ft} = 0.9756$, is defined as the FT for the test bearings in Fig. 8(b)–(d).

D. RUL Prediction

As presented in Section III-C, the estimated DEIs have been obtained from the trained CNN model. However, the DEIs of the test sets shown in Fig. 8(b)–(d) do not reach their fault limit, which need a regression algorithm to predict the estimated DEI till the end-of-life of each test bearing. An ϵ -SVR method is proposed to predict the upcoming degradation process of the test bearings. We conduct a training on the predicted DEI of bearing1_2. The sampling window size l and the moving size s in Fig. 3 is set to 50 and 1, respectively. The kernel function used in the prediction case is an RBF and penalty parameter C of the error term is chosen as 5.09 by grid search

TABLE III
PARAMETERS IN THE CNN MODEL

Layer	Filters	Kernel size/Stride	Output size
Input	1x2560x1
Conv1	64	1x100/50	1x50x64
Maxpooling1	...	1x2/2	1x25x64
Conv2	64	1x2/1	1x24x64
Maxpooling2	...	1x2/2	1x12x64
Flatten	1x768
FC1	1x100
Output	1x1

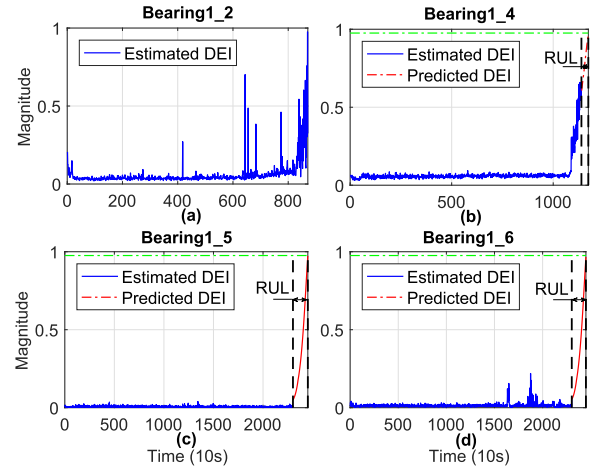


Fig. 9. (a) Estimated DEI of the bearing1_2. Predicted RULs of the test bearings are (b) 340 s for bearing1_4, (c) 1500 s for bearing1_5, and (d) 1480 s for bearing1_6. Green dotted lines in (b)–(d) represent the FTs.

and cross validation using the method of GridSearchCV [33]. We estimate the DEI after 1000 steps based on the existing DEI, and using the maximum value of predicted DEI of bearing1_2 (i.e., $L_{tf} = 0.9756$) to limit the termination time of the test sets.

Fig. 9(b)–(d) shows the predicted RULs $\hat{T}_{failure}$ of the test bearings till the failures occur. Red lines represent the predicted evolution of the bearings' degradation behavior using the ϵ -SVR method. The RULs are calculated as the difference between the final time when DEI reaches the FT and the time of the last known point of the test bearings. For bearing1_4, the predicted RUL is 340 s, while 1500 and 1480 s are the predicted RULs for bearing1_5 and bearing1_6, respectively.

TABLE IV
PROPOSED METHOD AND TESTED METHODS FOR COMPARISON

	DEI	CNN	SVR
C1	-	✓	✓
C2	✓	-	✓
Proposed method	✓	✓	✓

E. Comparison and Discussion

To assess the accuracy of the proposed method and compare with other existing approaches, two metrics are commonly adopted: 1) the relative percentage error ($Er\%$) which is given by (15) and 2) the exponential transformed accuracy (ETA) proposed in IEEE PHM 2012 [32]. ETA is an assessment index to distinguish the seriousness of the underestimate and overestimate of RUL prediction. It is clear that underestimate (i.e., early warning) is preferred over overestimation (i.e., warning after damage) to prevent more severe damage to the bearing. The formulas are expressed in (16) in the following:

$$Er = 100\% \times \frac{T_{\text{failure}} - \hat{T}_{\text{failure}}}{T_{\text{failure}}} \quad (15)$$

$$\text{ETA} = \begin{cases} \exp(-\ln(0.5) \frac{Er}{5}) & \text{if } Er \leq 0 \\ \exp(+\ln(0.5) \frac{Er}{20}) & \text{if } Er > 0 \end{cases} \quad (16)$$

where T_{failure} is the real RUL for the test bearing. A higher $|Er\%|$ means a worse RUL prediction result. On the other hand, ETA value varies from 0 to 1, and a higher score means a better RUL prediction result. In this article, Er is the common choice of most previous literature, thus it will be used for comparison.

In addition to the two metrics that evaluate the prediction performance for a specific bearing, three more assessment metrics, namely, average score S_{mean} , mean average error MAE, and the normalized root-mean-square error NRMSE, to make a comprehensive comparison of different methods, which are with the expression forms

$$S_{\text{mean}} = \frac{1}{\beta} \sum_{i=1}^{\beta} \text{ETA}_i \quad (17)$$

$$\text{MAE} = \frac{1}{\beta} \sum_{i=1}^{\beta} |T_{\text{failure}_i} - \hat{T}_{\text{failure}_i}| \quad (18)$$

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{\beta} \sum_{i=1}^{\beta} (T_{\text{failure}_i} - \hat{T}_{\text{failure}_i})^2}}{\frac{1}{\beta} \sum_{i=1}^{\beta} \hat{T}_{\text{failure}_i}} \quad (19)$$

where β is the number of test bearings.

To verify the benefits of the DEI and CNN techniques on RUL prediction, here, we also develop two other tested methods for comparison purpose (see Table IV). The proposed method and the tested methods are denoted and explained as follows.

- 1) *C1: CNN and ϵ -SVR*: This tested method uses a conventional linear time degradation label for CNN training rather than the nonlinear DEI. By this means, we can illustrate the impact of the DEI on RUL prediction.
- 2) *C2: DEI and ϵ -SVR*: Without training the CNN model for feature extraction, DEI in this tested method is extracted

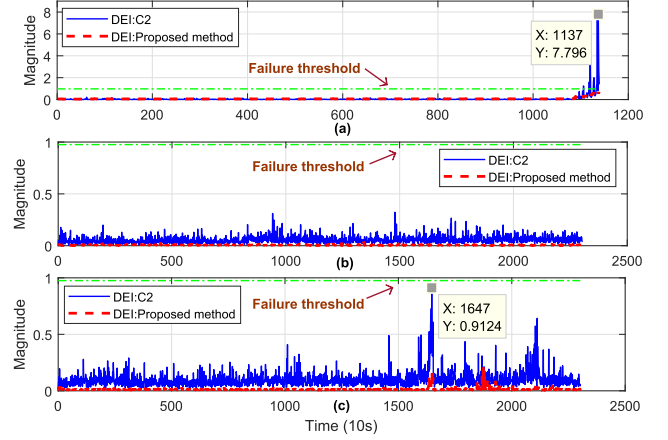


Fig. 10. Comparison of the estimated DEIs extracted using HHT (DEI:C2) and calculated by trained CNN model (DEI:Proposed method) of (a) test bearing1_4, (b) test bearing1_5, and (c) test bearing1_6.

manually and the ϵ -SVR is followed for the RUL prediction. Computing a DEI for a new bearing requires high computational power and longer time. In the meantime, the FT of each test bearing has to be predefined artificially, which increases the uncertainties of the RUL prediction affected by different working conditions. By this means, we can illustrate the impact of CNN modeling on the prediction of the final RUL.

- 3) *Proposed method: DEI-based CNN and ϵ -SVR*: This is the proposed framework which integrates DEI extraction, CNN, and ϵ -SVR into one framework.

The predicted numerical errors of the test bearings with the proposed approach and the tested methods are listed in Table V. Our approach achieves $Er\%$ of -0.29% , 6.83% , and -1.37% for bearing1_4, bearing1_5, and bearing1_6, respectively, which are much more smaller than the C1 and C2 methods. C1 uses a linear time degradation label for the training of the CNN model. The results show more than 19% prediction errors for test bearings and even 91.15% prediction error is obtained from bearing1_4, indicating that time degradation label is less effective than the DEI for the CNN training process. C2 is the method extracting the degradation indicator of the test bearings and defines the FTs manually. As testing bearing1_4, 1_5, and 1_6 operate under the same working conditions as bearing1_2, we employ the maximum and minimum value of bearing1_2 to normalize the extracted DEIs of C2 method. With the same working condition and same normalization parameters, FT of the proposed method could be reasonably used in C2 as well. To evaluate the impact of CNN modeling on the estimation of the final RUL, for the test bearings, the estimated DEIs extracted using HHT and calculated by trained CNN model are compared in Fig. 10. Without the CNN modeling procedure, one of the main drawback of the C2 method is that it requires a long time to calculate (~ 2 s of each sampling period). This limits the practical application of this method in the industry. Moreover, Fig. 10(a) shows that due to the uncertainties and huge amount of noise, the DEI extracted by C2 method has already exceeded the FT before the exact failure time, resulting in a 100% Er . Similarly, in Fig. 10(c), the extracted DEI of C2 method is much noised

TABLE V
COMPARISON OF PREDICTED RESULTS BETWEEN PROPOSED APPROACH AND OTHER TESTED METHODS

Methods	Bearing1_4 ($T_{failure} = 339$ s)			Bearing1_5 ($T_{failure} = 1610$ s)			Bearing1_6 ($T_{failure} = 1460$ s)		
	$\hat{T}_{failure}$	$E_r\%$	ETA	$\hat{T}_{failure}$	$E_r\%$	ETA	$\hat{T}_{failure}$	$E_r\%$	ETA
Proposed approach	340 s	-0.29%	0.96	1500 s	6.83%	0.79	1480 s	-1.37%	0.83
C1	30 s	91.15%	0.04	820 s	49.07%	0.18	1181 s	19.11%	0.52
C2	N/A (0 s)	N/A (100%)	0.03	1140 s	29.19%	0.36	1080 s	26.02%	0.41

TABLE VI
COMPARISON OF PREDICTED RESULTS BETWEEN PROPOSED APPROACH AND EXISTING MAJOR APPROACHES

Methods	$E_r\%$					S_{mean}	MAE	NRMSE
	Bearing1_4	Bearing1_5	Bearing1_6	Bearing2_4	Bearing2_6			
Proposed approach	-0.29%	6.83%	-1.37%	5.75%	1.55%	0.87	46.2	0.05
FCAMN [23]	21.95%	-15.22%	-5.74%	-	-	0.35	696.7	0.13
Multi-scale CNN [24]	10.69%	-148.20%	-21.51%	-	-	0.25	1003.3	0.50
CWT-CNN [34]	20.35%	11.18%	34.93%	-1.44%	-42.64%	0.46	265.8	0.29
RNN [35]	62.07%	-22.98%	21.23%	-19.42%	-13.95%	0.17	586.0	0.57
Particle Filtering [36]	5.60%	100.00%	28.08%	8.63%	58.91%	0.42	583.8	1.29
FFT+Ratio [37]	80.00%	9.00%	-5.00%	10%	49%	0.44	252	0.32
LSTM [38]	38.69%	-99.40%	-120.07%	19.81%	17.87%	0.26	996.2	0.57
SOM [39]	-20.94%	-278.26%	19.18%	51.80%	-20.93%	0.16	1164.2	1.02

than that of the proposed method. At 16 470 s, the magnitude of DEI:C2 is almost close to the FT, this phenomenon might lead to a waste of sources due to excessive underestimation of RUL. The comparison results in **Table V** demonstrate the benefits of using DEI and CNN in estimating RULs of REBs.

To demonstrate the generality of our proposed RUL estimation framework, other test bearings (i.e., bearing2_4 and bearing2_6) that operate undergoing a different external load and rotational speed are analyzed. Bearing2_2 is used for training, and the trained CNN model is obtained without changing any hyperparameters and the architecture of the model in **Table III**. We just fine-tune the ϵ -SVR forecasting model by changing the penalty parameter C from 5.09 to 7.09, resulting in 5.75% and 1.55% E_r for bearing2_4 and bearing2_6, respectively. Good RUL prediction results of bearings with different operating conditions indicate the repeatability and robustness of our proposed method, with respect to the hyperparameters and the architecture of the CNN.

To further verify the proposed approach, the predicted numerical errors of RULs generated by the proposed method and eight published methods are compared and listed in **Table VI**. Other published approaches include a recurrent neural network method-based health indicator [35], the method proposed by the winner of the IEEE PHM 2012 prognostic [37], a convolutional long-short-term memory network (LSTM) method [38], and a self-organization mapping (SOM) method, etc.. In addition, recent CNN-based approaches, including the frozen convolution and activated memory network (FCAMN) [23], multiscale CNN [24], and continuous wavelet transform CNN (CWT-CNN) [34] are also compared. The results of the comparison shown in **Table VI** confirm that our approach significantly outperforms the referenced methods with the highest average score $S_{mean} = 0.87$, and the smallest MAE = 46.2 and NRMSE = 0.05. In particular, a -0.29% E_r of bearing1_4 is achieved, owing to a 1-s absolute time error. This result benefits from a good nonlinear degradation indicator extracted using the HHT method. In addition, the CNN is a powerful tool

for discovering the hidden pattern of the extracted degradation indicator and the underlying bearing system, further increasing the accuracy of the predicted RUL.

It can be concluded from the experimental results that the proposed data-driven RUL estimation approach has much smaller prediction errors, compared with both the tested methods in this article and the other methods published in previous studies.

IV. CONCLUSION

In this article, a data-driven framework for RUL prediction of REB is presented using the HHT method, a CNN model, and an ϵ -SVR forecasting model. A nonlinear degradation indicator DEI is first extracted from the raw vibration signals using the HHT method, which is defined as the label for the training. A CNN model is trained to discover the hidden pattern between the extracted DEI and the raw vibration data of the training bearing. In this way, predicted DEIs are automatically obtained when applying the trained CNN model to the test bearings. Finally, the RULs of the testing bearings are obtained using an ϵ -SVR forecasting model. An experimentation platform that allows to observe the accelerated degradation process of bearings is employed to validate the proposed framework. The proposed framework achieves much smaller prediction errors for RUL predictions than previous published approaches.

Future work includes the application of the proposed framework to a wider range of case studies on experimental data in other applications [40] and the investigation of other potential degradation labels to achieve even higher accuracy in estimating RUL.

REFERENCES

- [1] T. A. Harris, *Rolling Bearing Analysis*. Hoboken, NJ, USA: Wiley, 2001.
- [2] Y. Yuan, H.-T. Zhang, Y. Wu, T. Zhu, and H. Ding, "Bayesian learning-based model-predictive vibration control for thin-walled workpiece machining processes," *IEEE/ASME Trans. Mechatronics*, vol. 22, no. 1, pp. 509–520, Feb. 2017.

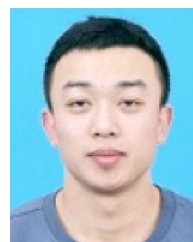
- [3] A. Heng, S. Zhang, A. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mech. Syst. Signal Process.*, vol. 23, no. 3, pp. 724–739, 2009.
- [4] L. R. Rodrigues, "Remaining useful life prediction for multiple-component systems based on a system-level performance indicator," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 141–150, Feb. 2018.
- [5] I. V. de Bessa, R. M. Palhares, M. F. S. V. D'Angelo, and J. E. C. Filho, "Data-driven fault detection and isolation scheme for a wind turbine benchmark," *Renewable Energy*, vol. 87, pp. 634–645, 2016.
- [6] Z.-Q. Wang, C.-H. Hu, and H.-D. Fan, "Real-time remaining useful life prediction for a nonlinear degrading system in service: Application to bearing data," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 211–222, Feb. 2018.
- [7] Y. Li, T. Kurfess, and S. Liang, "Stochastic prognostics for rolling element bearings," *Mech. Syst. Signal Process.*, vol. 14, no. 5, pp. 747–762, 2000.
- [8] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation—A review on the statistical data driven approaches," *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 1–14, 2011.
- [9] J. Wang, P. Fu, L. Zhang, R. X. Gao, and R. Zhao, "Multi-level information fusion for induction motor fault diagnosis," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 5, pp. 2139–2150, Oct. 2019.
- [10] H. Gao, L. Liang, X. Chen, and G. Xu, "Feature extraction and recognition for rolling element bearing fault utilizing short-time fourier transform and non-negative matrix factorization," *Chin. J. Mech. Eng.*, vol. 28, no. 1, pp. 96–105, 2015.
- [11] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Process.*, vol. 96, pp. 1–15, 2014.
- [12] Q. Meng and L. Qu, "Rotating machinery fault diagnosis using Wigner distribution," *Mech. Syst. Signal Process.*, vol. 5, no. 3, pp. 155–166, 1991.
- [13] A. Soualhi, K. Medjaher, and N. Zerhouni, "Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 52–62, Jan. 2015.
- [14] J. Wu, C. Wu, S. Cao, S. W. Or, C. Deng, and X. Shao, "Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 1, pp. 529–539, Jan. 2019.
- [15] G. Cheng, Y.-L. Cheng, L.-H. Shen, J.-B. Qiu, and S. Zhang, "Gear fault identification based on Hilbert–Huang transform and SOM neural network," *Measurement*, vol. 46, no. 3, pp. 1137–1146, 2013.
- [16] Z. Tian, "An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring," *J. Intell. Manuf.*, vol. 23, no. 2, pp. 227–237, 2012.
- [17] W. Wang, "An adaptive predictor for dynamic system forecasting," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 809–823, 2007.
- [18] J. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1803–1836, 2011.
- [19] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
- [20] M. Sun, I. Konstantelos, and G. Strbac, "A deep learning-based feature extraction framework for system security assessment," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5007–5020, Sep. 2019.
- [21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [22] Y. Yuan *et al.*, "A general end-to-end diagnosis framework for manufacturing systems," *Nat. Sci. Rev.*, Nov. 2019, doi: [10.1093/nsr/nwz190](https://doi.org/10.1093/nsr/nwz190).
- [23] Z. Chen, X. Tu, Y. Hu, and F. Li, "Real-time bearing remaining useful life estimation based on the frozen convolutional and activated memory neural network," *IEEE Access*, vol. 7, pp. 96583–96593, 2019.
- [24] J. Zhu, N. Chen, and W. Peng, "Estimation of bearing remaining useful life based on multiscale convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3208–3216, Apr. 2019.
- [25] L. Ren, Y. Sun, H. Wang, and L. Zhang, "Prediction of bearing remaining useful life with deep convolution neural network," *IEEE Access*, vol. 6, pp. 13041–13049, 2018.
- [26] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, "Using Bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 188–201, Jan. 2020.
- [27] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, 2018.
- [28] N. Tandon and A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribology Int.*, vol. 32, no. 8, pp. 469–480, 1999.
- [29] B. Neyshabur, S. Bhojanapalli, D. McAllester, and N. Srebro, "Exploring generalization in deep learning," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, 2017, pp. 5947–5956.
- [30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CORR*, vol. abs/1412.6980, 2014.
- [31] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, "Remaining useful life estimation based on nonlinear feature reduction and support vector regression," *Eng. Appl. Artif. Intell.*, vol. 26, no. 7, pp. 1751–1760, 2013.
- [32] P. Nectoux *et al.*, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in *Proc. IEEE Int. Conf. Prognostics Health Manage.*, 2012, pp. 1–8.
- [33] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [34] Y. Yoo and J.-G. Baek, "A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network," *Appl. Sci.*, vol. 8, no. 7, 2018, Art. no. 1102.
- [35] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, 2017.
- [36] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, and J. Dybala, "A model-based method for remaining useful life prediction of machinery," *IEEE Trans. Rel.*, vol. 65, no. 3, pp. 1314–1326, Sep. 2016.
- [37] E. Sutrisno, H. Oh, A. S. S. Vasan, and M. Pecht, "Estimation of remaining useful life of ball bearings using data driven methodologies," in *Proc. IEEE Conf. Prognostics Health Manage.*, 2012, pp. 1–7.
- [38] A. Z. Hinchí and M. Tkouat, "Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network," *Procedia Comput. Sci.*, vol. 127, pp. 123–132, 2018.
- [39] S. Hong, Z. Zhou, E. Zio, and K. Hong, "Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method," *Digit. Signal Process.*, vol. 27, pp. 159–166, 2014.
- [40] Y. Yuan *et al.*, "Data driven discovery of cyber physical systems," *Nature Commun.*, vol. 10, no. 1, pp. 1–9, 2019.



deep learning applications.

Cheng Cheng received the B.Eng. degree in measurement, control technology, and instrument from Tianjin University, Tianjin, China, in 2012 and the M.Sc. and Ph.D. degrees in control systems from Imperial College London, London, U.K., in 2013 and 2018, respectively.

Since 2018, she has been a Postdoctoral Researcher with the Huazhong University of Science and Technology, Wuhan, China. Her research interests include robust control, mechatronic systems modeling and simulation, and



Guijun Ma received the bachelor's degree, in 2017, from the Huazhong University of Science and Technology, Wuhan, China, where he is currently working toward the Ph.D. degree in mechanical engineering with the School of Mechanical Science and Engineering.

His research interests include machine fault diagnosis and remaining useful life prediction.



Yong Zhang received the B.Sc. degree in mathematics from Jiangsu Normal University, Xuzhou, China, in 2001, the M.Sc. degree in applied mathematics from Three Gorges University, Yichang, China, in 2007, and the Ph.D. degree in control theory and control engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2010.

From 2014 to 2015, he was a Visiting Scholar with the Department of Information Systems and Computing, Brunel University London, Uxbridge, U.K. He is currently an Associate Professor with the Department of Automation, School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan, China. He has authored more than ten papers in refereed international journals. He is a very active reviewer for many international journals. His current research interests include remaining useful life prediction of key equipment and fault diagnosis and fault tolerant control of networked systems.



Mingyang Sun (Member, IEEE) received the Ph.D. degree in control science and engineering from the Department of Electrical and Electronic Engineering, Imperial College London, U.K., in 2017.

From 2017 to 2019, he was a Research Associate and a DSI Affiliate Fellow with Imperial College London. He is currently a Professor (Tenure-Track) under the Hundred Talents Program with the College of Control Science and Engineering, Zhejiang University, Hangzhou, China. He is also a Visiting Researcher with Imperial College London. His research interests include artificial intelligence in energy systems and cyber-physical energy system security and control.



Fei Teng (Member, IEEE) received the B.Eng. degree in electrical engineering from Beihang University, Beijing, China, in 2009, and the Ph.D. degree in electrical engineering from Imperial College London, London, U.K., in 2015.

He is currently a Lecturer with the Department of Electrical and Electronic Engineering, Imperial College London. His research interests include power system operation, cyber-physical system modeling, and data analytics.



Han Ding (Senior Member, IEEE) received the Ph.D. degree in mechanical engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 1989.

From 1993 to 1994, he was with the University of Stuttgart, Stuttgart, Germany, supported by the Alexander von Humboldt Foundation. Since 1997, he has been a Professor with HUST, where he is currently the Director of the State Key Lab of Digital Manufacturing Equipment and Technology. From 2001 to 2006, he was a Cheung Kong Chair Professor with Shanghai Jiao Tong University, Shanghai, China. His research interests include robotics, multiaxis machining, and control engineering.

Dr. Ding was an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING from 2004 to 2007. He is an Editor for the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING and a Senior Editor for the IEEE ROBOTICS AND AUTOMATION LETTERS. He was elected as a Member of the Chinese Academy of Sciences, in 2013.



Ye Yuan (Member, IEEE) received the B.Eng. degree (Valedictorian) from the Department of Automation, Shanghai Jiao Tong University, Shanghai, China, in 2008, and the M.Phil. and Ph.D. degrees from the Department of Engineering, University of Cambridge, Cambridge, U.K., in 2009 and 2012, respectively, all in control theory.

Since 2016, he has been a Full Professor with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China. Prior to this, he was a Postdoctoral Researcher with the University of California Berkeley, Berkeley, CA, USA and a Junior Research Fellow with Darwin College, University of Cambridge. His research interests include system identification and control with applications to cyber-physical systems.

Dr. Yuan was the recipient of the China National Recruitment Program of 1000 Talented Young Scholars, the Dorothy Hodgkin Postgraduate Awards, Microsoft Research Ph.D. Scholarship, and the Best of the Best Paper Award at the IEEE Power and Energy Society General Meeting.