

Degradation-Aware Remaining Useful Life Prediction With LSTM Autoencoder

Ji-Yan Wu^{id}, Min Wu^{id}, *Senior Member, IEEE*, Zhenghua Chen^{id},
Xiao-Li Li^{id}, *Senior Member, IEEE*, and Ruqiang Yan^{id}

Abstract—The remaining useful life (RUL) prediction plays a pivotal role in the predictive maintenance of industrial manufacturing systems. However, one major problem with the existing RUL estimation algorithms is the assumption of a single health degradation trend for different machine health stages. To improve the RUL prediction accuracy with various degradation trends, this article proposes an algorithm dubbed degradation-aware long short-term memory (LSTM) autoencoder (AE) (DELTA). First, the Hilbert transform is adopted to evaluate the degradation stage and factor with the real-time sensory signal. Second, we adopt LSTM AE to predict RUL based on multisensor time-series data and the degradation factor. Distinct from the existing studies, the proposed framework is able to dynamically model the degradation factor and explore latent variables to improve RUL prediction accuracy. The performance of DELTA is evaluated with the open-source FEMTO bearing data set. Compared with the existing algorithms, DELTA achieves appreciable improvements in the RUL prediction accuracy.

Index Terms—Industrial internet-of-things (IIoT), long short-term memory (LSTM) autoencoder (AE), prognostic technique, remaining useful life (RUL).

I. INTRODUCTION

THE rapid development of the Industrial Internet of Things (IIoT) promotes the technological evolutions in manufacturing systems. In the IIoT environment, there are various sensors used for monitoring and controlling the industrial production processes. These heterogeneous sensor readings are able to reflect the healthy levels of machines/components. Therefore, it is desirable to take advantage of these sensory data to perform condition-based maintenance rather than the traditional time-based maintenance [1], [2]. Such prognostic technique can be achieved by developing algorithms for remaining useful life (RUL) prediction [3]–[5].

In particular, the practical challenges for RUL prediction are as follows.

Manuscript received October 1, 2020; revised December 25, 2020; accepted January 17, 2021. Date of publication February 1, 2021; date of current version March 3, 2021. This work was supported in part by the A*STAR Industrial Internet of Things Research Program under Grant RIE2020 IAF-PP and Grant A1788a0023 and in part by the National Natural Science Foundation of China under Grant 51835009. The Associate Editor coordinating the review process was Loredana Cristaldi. (*Corresponding author: Min Wu.*)

Ji-Yan Wu, Min Wu, Zhenghua Chen, and Xiao-Li Li are with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 138632 (e-mail: wu_jiyan@i2r.a-star.edu.sg; wumin@i2r.a-star.edu.sg; chen_zhenghua@i2r.a-star.edu.sg; xlli@i2r.a-star.edu.sg).

Ruqiang Yan is with the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: yanruqiang@xjtu.edu.cn).

Digital Object Identifier 10.1109/TIM.2021.3055788

- 1) *Health Degradation Trend*: In complex machines with various components, it is extremely difficult or even impossible to develop mechanical-condition-based models for degradation trend analysis. For instance, the data-driven RUL prediction algorithms assume the exponential degradation trend. These algorithms are applicable when the machine health measurements are unavailable. However, the exponential degradation trend assumption may not stand in scenarios, such as approaching failure.
- 2) *Noisy Sensory Readings*: Sensory readings are often interfered with by different levels of environmental noise. The noisy levels of various sensors may even differ from each other significantly.
- 3) *Temporal Dependencies*: The various machinery components interact with each other in a complex way, and this results in complicated temporal dependencies among different sensors. It is important and necessary to exploit the sensor readings to capture the complex operational behaviors of machines with regard to the temporal dependencies.

To address the above issues, there are a plethora of RUL prediction schemes proposed in recent studies that can be classified into the modeling- [1]–[4], [6] and data-based [7]–[12] algorithms. However, the data-based solutions require a large amount of training data to develop the highly accurate model. Conventionally, a system degradation model is assumed in the modeling-based approaches for the specific component without sufficient evidence/proof [4].

This article overcomes the practical challenges by adopting long short-term memory (LSTM) autoencoder (AE) for the RUL prediction. The main motivation for us to use deep learning in RUL prediction is the complexity of different bearing systems. The degradation patterns and features of various bearing systems significantly vary from each other because of the differences in physical properties and operational environments. The traditional statistical model-based techniques are normally developed for specific bearing systems based on mathematical modeling and expert knowledge. The main limitation of such modeling-driven techniques is the applicability to other bearing systems. Deep learning algorithms are able to automatically learn the degradation features/patterns of different bearing systems from historical sensory data. Therefore, we can develop RUL prediction solutions for various bearing systems by

retraining the deep learning model with different sensory data.

The mechanism of our proposed approach can be described as follows: an LSTM-based encoder is adapted to convert a multivariate input sequence to a fixed-dimension data vector. Then, an LSTM-based decoder adopts this vector representation to generate a target sequence. LSTM AE-based solutions learn a model to rebuild the normal data such that the generated model is able to rebuild the subsequences which reflect normal behavior. One main problem with the existing algorithms [7]–[9] is the assumption of degradation trends. To remedy this disadvantage, this article develops a degradation trend evaluation algorithm based on the Hilbert transform [13].

The proposed degradation-aware LSTM AE (DELTA) is able to capture different levels of RUL degradation trends and, thus, minimize the possibility of early/late predictions. Our experimental results indicate that the degradation factor increases as the machine condition deteriorates. Furthermore, our proposed DELTA is not dependent on expert domain knowledge or degradation trend assumption. In particular, our contributions in this research can be summarized as follows.

- 1) We propose a degradation-aware RUL prediction scheme that effectively integrates the following two key modules: 1) a healthy stage and degradation factor evaluation algorithm that leverages the Hilbert transform to identify the current machine status and 2) an RUL prediction algorithm with LSTM AE by means of latent variable exploration and degradation similarity matching.
- 2) We evaluate the performance with the open-source FEMTO data set [14] for bearing system degradation. Experimental results demonstrate the proposed DELTA framework achieves higher prediction accuracy than multilayer perceptron (MLP), extreme gradient boost (XGBoost), random forest regression (RFR), convolutional neural network (CNN) [5], LSTM, and support vector regression (SVR) [4], respectively.

The remainder of this article includes the following parts. Section II briefly reviews the related work close to this article. In Section III, we present the system framework and introduce the proposed algorithms in detail. The performance evaluation is provided in Section IV.

II. RELATED WORK

In this section, we review the recent studies on RUL estimation that can be generally divided into: 1) data-driven; 2) model-driven; and 3) hybrid methods.

A. Data-Driven RUL Prediction

In [3], the extended Kalman filter is adopted for the RUL estimation. In [4], the relationship between health indicator and sensory data values is directly modeled using SVR. Furthermore, an off-line wrapper variable is used for selection before model training. Babu *et al.* [5] adopt deep CNN based on regression scheme for RUL prediction. The convolutional and pooling filters in [5] are used in conjunction with the temporal dimension on the sensory data captured from multiple

channels. Mosallam *et al.* [15] propose a data-driven method for RUL prediction with off-line and online phases for health indicator and RUL predictor calculations. Orchard *et al.* [16] develop an RUL prediction algorithm that merges the different information obtained from the multiple features and appropriately correlates with the machinery failure. Then, the authors estimate the RUL using a particle filtering-based algorithm. Miao *et al.* [17] combine an unscented particle filter scheme and degradation modeling to estimate the RUL. The probability distributions of the last step for the online update are considered as the RUL value distribution. The belief functions are adopted by Ramasso *et al.* [18] to categorize the original sensory data into four healthy states. Then, the RUL is estimated as the duration from degradation to the failure state.

In [19], a joint fuzzy mean clustering and neural network method is adopted to classify the sensory data into dynamic health status. The RUL is calculated based on the time stamp of the failure status. Miaot *et al.* [31] design a dual-task deep long short-term memory networks for joint degradation evaluation and RUL prediction. This improves the robustness and accuracy of the assessment and prediction results. Xia *et al.* [32] propose a two-stage RUL prediction approach with an AE-based DNN and regression models based on shallow neural networks.

B. Model-Driven RUL Prediction

In instance-based learning research, the system degradation is expressed as a health indicator obtained with the fuzzy model [20], unsupervised kernel regression modeling [21], and linear regression [22], or using upper and lower envelopes of a moving path to compute polygons [23]. Then, the RUL is calculated as a weighted smooth moving average of all useful life values from similar cases. The particle swarm optimization is adopted by Qin *et al.* [24] to obtain SVR configuration parameters. Loutas *et al.* [27] use the Wiener entropy to monitor conditions and locate critical system faults. Lei *et al.* [29] propose a machine RUL prediction method that includes: 1) a stochastic model that characterizes the machine degradation process with multiple variable sources and 2) a Kalman particle filtering scheme to predict RUL and estimate the system states.

C. Hybrid RUL Prediction

There are also some studies using pattern matching for prognostic. These models first estimate the variation of a failure signal, and the RUL is then estimated as the duration for the signal to approach the end of life (EOL). These approaches generally require adjusting failure threshold values. The SVR [24] or support vector machines [25] are adopted to capture degradation models. The minimal value obtained with the three mathematical models in reference [25] is used to estimate the RUL. In [26], the RUL prediction is conducted only for the state near the EOL and is defined as a percentage of EOL values. With regard to the threshold value of the modeled critical operational condition, the RUL is then predicted. Dong *et al.* [28] integrate SVR-PF model to predict

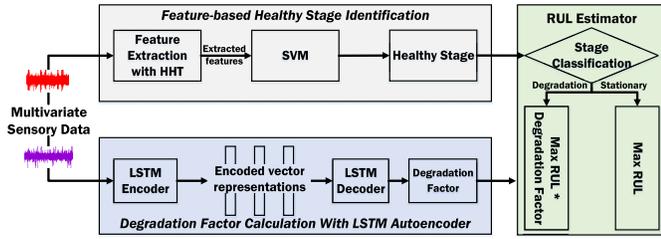


Fig. 1. System design of the proposed DELTA framework.

the component degradation time. Then, the RUL is estimated by profiling the probability distribution.

In summary, a failure threshold is generally required in degradation and regression-based prognostic methods. However, a static threshold is difficult to estimate, and such static parameters are not adaptive to environmental changes. On the other side, the pattern matching methods normally have no dependence on threshold values. RUL prediction is the procedure of tracing the correct trend following its EOL value. Nonetheless, this approach cannot derive the right solutions in the case that matched patterns are absent. This article leverages the advantages of machine learning-based classification and regression. The special features of different stages are leveraged to improve the prediction accuracy.

III. DEGRADATION-AWARE RUL PREDICTION

The system diagram of the proposed DELTA is presented in Fig. 1. The multivariate real-time sensory data are collected from various sensors in manufacturing systems. With the heterogeneous sensory data, DELTA performs RUL prediction with the three steps: 1) feature-based health stage identification; 2) degradation factor calculation with LSTM AE; and 3) RUL estimation with the health stage and degradation factor. The main contributions to adopt the LSTM AE for our considered RUL prediction problem include: 1) interleaved usage of LSTM unit and encoder–decoder pairs for degradation factor estimation; 2) introduce a linear layer with bias vector above the decoder to calculate the updated degradation factor; and 3) normalize the reconstruction errors from multiple sequences for the final prediction value. The following subsections will introduce the details of each step.

A. Feature-Based Healthy Stage Identification

The objective of the health stage identification is to estimate the current health level. Many feature-based algorithms can be adopted for this healthy stage identification. In this article, the Hilbert–Huang transform (HHT) is employed for the degradation parameter estimation.

HHT is a signal processing algorithm developed for time–frequency data to analyze nonstationary features. In recent studies, HHT is widely used in signal/data processing, including prognostic and fault detection [13]. Specifically, a time-series signal $x(t)$ is divided by HHT into several intrinsic modal functions (IMFs) that stand for the mean signal. These IMFs are derived from empirical mode decomposition (EMD), and the input signal is expressed in different frequency bands.

Suppose that the notation m_{i0} denote the average value of the lower and upper envelopes for the vibration signal $x(t)$. The first important component denoted with h_{i0} is the difference between the signal $x(t)$ and m_{ik} : $x(t) - m_{ik}(t) = h_{ik}(t)$. h_{ik} is classified as an IMF if it satisfies the conditions as follows.

- 1) In the time-series signal $x(t)$, the number of zero-crossings and the number of extremes should either equal to each other or differ at most by one.
- 2) For each time step t , the average value of the locally maximal envelope and the locally minimal envelope value is close to zero.

This shifting process is running iteratively until time k on h_{ik} until the mean value between the upper and lower envelopes is close to zero at each point

$$h_{ik}(t) = h_{i(k-1)}(t) - m_{i(k-1)}(t) \quad (1)$$

in which $m_{i(k-1)}(t)$ represents the mean value and k is the iteration number.

The signal $x(t)$ is converted into multiple IMFs with EMD. The first IMF is the highest frequency signal component. Then, the low-frequency signal components are decomposed the other IMFs. The frequency band of each IMF differs from other functions with regard to the signal value $x(t)$. The identification process includes extracting the instantaneous amplitude $a_i(t)$, the Hilbert marginal spectrum $h_i(f)$ of the frequency domain, and the instantaneous frequency $f_i(t)$. In the following analysis, we use the symbol $c_i(t)$ to represent a modal function. The mathematical expression of an IMF $c_i^A(t)$ is presented with the following equation:

$$c_i^A(t) = c_i(t) + j \cdot c_i^H(t) = a_i(t)e^{j\theta_i(t)} \quad (2)$$

in which $c_i^H(t)$ denotes the Hilbert transform of $c_i(t)$, that is,

$$c_i^H(t) = \frac{1}{\pi} P \int \frac{c_i(s)}{t-s} ds \quad (3)$$

in which P stands for the Cauchy principal value.

Based on the analytical IMF $c_i(t)$, the phase $\theta_i(t)$ and instantaneous amplitude $a_i(t)$ can be obtained from the polar coordinate. The two values can be expressed as follows:

$$\begin{cases} a_i(t) = \sqrt{c_i^2 + c_i^{H2}} \\ \theta_i(t) = \tan^{-1}\left(\frac{c_i^H}{c_i}\right). \end{cases} \quad (4)$$

The instantaneous frequency $f_i(t)$ can be calculated from the phase $\theta_i(t)$ with the equation of

$$f_i(t) = \frac{1}{2\pi} \frac{d\theta_i(t)}{dt}. \quad (5)$$

The formula to calculate the Hilbert marginal spectrum is

$$h_i(f) = \int h_i(f, t) dt = \int a_i^2(f_i, t) dt \quad (6)$$

in which $h_i(f, t)$ denotes the Hilbert spectral density calculated from the i th modal function of the input signal $x(t)$. The parameter $a_i(f_i, t)$ integrates the amplitude $a_i(t)$ and the instantaneous frequency $f_i(t)$.

After the identification of the frequency bands, it is possible to pick an IMF based on the different frequencies of various

faults. These frequencies can be extracted according to different types of sensory data (e.g., vibration, acoustic, voltage, and current signal) measured in practical manufacturing systems. For instance, the IMF is based on the maximum value of the inner race frequency f_{ir} , the outer race frequency f_{or} , and the ball frequency f_b for the bearing systems

$$c_i(t) \rightarrow \max_{1 \leq i \leq n} [h_i(f_{or}, f_{ir}, f_b)]. \quad (7)$$

These frequencies depend on the different physical properties of the machine/component.

Totally, the three indicators are extracted with the filtered IMFs: $h_i(f_{or})$, $h_i(f_{ir})$, and $h_i(f_b)$. These indicators are involved in the diagnosis of health degradation. The identification of system degradation can be understood as status classification problem. The objective is to obtain a status expressed with symbol $x = [h_j(f_{or})h_j(f_{ir})h_j(f_b)]$ for time step t in one of the classes that represent the degradation stages ($\Omega_1, \Omega_2, \dots, \Omega_M$).

The M -classes regroup historical observations collected from components/machines with the same features and the same operational environments of the testing samples. These stages are identified with the SVM technique, which leverages the historical sensory data as training samples to develop a classifier $f(x)$. Suppose that there is a training data set $(x_i, \Omega_i)_{i \in 1,2,\dots,l}$, $x_i \in \mathbb{R}^3$, $\Omega_i \in \{-1, +1\}$. -1 and 1 represent the healthy and degradation stages, respectively. The SVM classifier $f(x)$ is defined as follows:

$$f(x) = \text{sign} \left[\sum_{i=1}^l \alpha_i \Omega_i K(x_i, x) + b \right] \quad (8)$$

in which Ω_i is selected in one of two values: 1 or -1 . b is a real constant, l represents the number of elements in the training samples, α_i denotes the Lagrange multipliers, and $K(\cdot, \cdot)$ stands for the kernel function.

As a nonlinear hyperplane, the SVM classifier $f(x)$ for healthy stage classification is expressed as

$$\begin{cases} w^T \cdot \Phi(x) + b \geq 1 \\ w^T \cdot \Phi(x) + b \leq -1 \end{cases} \quad (9)$$

which can also be stated as follows:

$$\Omega_i [w^T \cdot \Phi(x) + b] \geq 1, \quad \Omega_i \in \{-1, +1\}, i \in \{1, 2, \dots, l\}. \quad (10)$$

In the above equation, w represents the hyperplane margin. $\Phi: x \rightarrow \Phi(x)$ represents a nonlinear function which maps the time-series value x into a higher dimensional space. Dot products with w for classification can be operated with the following kernel method:

$$w^T \cdot \Phi(x) = \sum_{i=1}^l \alpha_i \Omega_i \Phi(x_i) \cdot \Phi(x). \quad (11)$$

The relationship between the kernel and the transform $\Phi(x_i)$ is expressed as

$$K(x_i, x) = \Phi(x_i) \cdot \Phi(x) \quad (12)$$

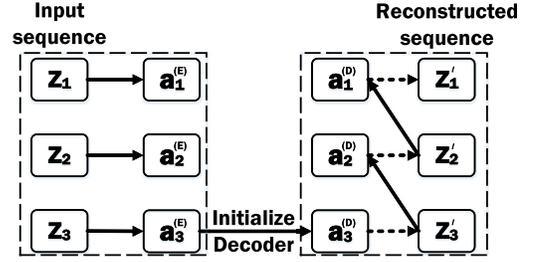


Fig. 2. LSTM encoder–decoder structure.

in which $K(x_i, x)$ denotes the form, $K(x_i, x) = \exp\{-\|x - x_i\|^2/\sigma^2\}$ radial basis function (RBF), degree polynomial $d: K(x_i, x) = (x_i^T \cdot x + 1)^d$, and hyperbolic tangent $K(x_i, x) = \tanh(\beta x_i \cdot x + c)$.

B. Degradation Factor Estimation With LSTM Autoencoder

The main motivation for us to introduce the degradation factor estimation stems from the different degradation trends. Depending on various factors, the bearing degradation may be different for distinct bearings. Assuming that no other information is available about the other components than the rotating system and that load and speed are constant, one has to use only the data collected via the sensors located around the bearings. Moreover, nothing is known about the nature and the origin of the degradation (balls, inner or outer races, cage, and so on); therefore, data-driven techniques have to be applied. According to the bearing and to the way the degradation evolves, the fault modes can be slightly different for distinct bearings. In the gradual degradation cases, the degradation factor gradually increases until the end of the useful lifetime. For other sudden degradation cases, the degradation factor sharply decreases.

The LSTM encoder–decoder structure is presented in Fig. 2. To adopt LSTM AE for the RUL prediction, the sensory data for the lifetime of multiple instances are available for training purposes. An LSTM unit is a recurrent module that leverages the memory cell activation c_{t-1} , input data z_t , and the hidden state activation to calculate the hidden state activation a_t at a time t . This LSTM basic unit combines the memory cell c and three types of gates, i.e., forget gate f , input gate i , output gate o , to determine the necessity of remembering the input.

Let the notation $T_{n1,n2}$ denote an affine transform of the form $z \rightarrow Wz + b$ for vector b and matrix W . Based on the current input z_t , the values for input gate i , output gate o , forget gate f , hidden state a , and cell activation c at time t can be calculated. The memory cell value c_{t-1} and previous hidden state a_{t-1} are calculated with the following formula:

$$\begin{bmatrix} i_t \\ f_t \\ o_t \\ g_t \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh h \end{bmatrix} T_{m+n,4n} \begin{bmatrix} z_t \\ a_{t-1} \end{bmatrix}.$$

In the above equation, $\sigma(z) = (1/(1 + e^{-z}))$, and $\tanh h(z) = 2\sigma(2z) - 1$. The operations σ and $\tanh h$ are applied elementwise

$$c_t = f_t \cdot c_{t-1} + i_t \cdot g_t.$$

We consider the sliding windows to construct $(L - l + 1)$ subsequences for an L -cycle training sample. The LSTM AE is trained to reconstruct the l -length subsequences extracted from all the collected samples. From the input time-series data, the LSTM encoder learns a fixed-length vector format. The LSTM decoder adopts this representation to rebuild the time series using the previously predicted value and the current hidden state. c denotes the number of LSTM units in the hidden layer of the encoder. For the time series $Z = [z_1, z_2, \dots, z_l]$ data, a_t denotes the encoder hidden state at time t .

The value x_t and the hidden state $a_{t-1}^{(E)}$ of the encoder at time $t - 1$ are employed to predict the encoder hidden state $a_t^{(E)}$ at time t . If $a_t^{(D)} = a_t^{(E)}$, the encoder hidden state $a_t^{(E)}$ at the end of the input sequence is adopted as the decoder initial state $a_1^{(D)}$. A linear layer with bias vector $b \in R^m$ and weight matrix $w \in R^{c \times m}$ above the decoder is used to calculate the value of $z'_t = w^T a_t^{(D)} + b$. In the training process, z_t is the input to estimate the state $a_{t-1}^{(D)}$ and z'_{t-1} mapped to the parameter z_{t-1} . The predicted value z'_t is input to the decoder to estimate z'_{t-1} and $a_{t-1}^{(D)}$ for the inference. For any data point $z_t^{(u)}$ in sample u , the reconstruction error $e_t^{(u)}$ is estimated with the following formula:

$$e_t^{(u)} = \|z_t^{(u)} - z'_t{}^{(u)}\|. \quad (13)$$

The model is trained to minimize the objective value $E = \sum_{u \in U} \sum_{t=1}^l (e_t^{(u)})^2$, in which U denotes the data sample set.

A point z_t in a time-series Z is estimated with multiple subsequences $Z(j, l)$ in which $j = t - l + 1, t - l + 2, \dots, t$. Therefore, z_t is a part of multiple overlapping subsequences. Consequently, each point in the original time series involved in the train instance is predicted as the number of belonged subsequences. A mean value for all the estimations at a time step is recorded as the final predicted value. The difference in actual and predicted values for a point is used as a nonnormalized health degradation factor on that time step.

The estimation error $e_t^{(u)}$ is normalized to obtain the target health degradation factor $h_t^{(u)}$ as

$$h_t^{(u)} = \frac{e_M^{(u)} - e_t^{(u)}}{e_M^{(u)} - e_m^{(u)}} \quad (14)$$

in which $e_m^{(u)}$ and $e_M^{(u)}$ represent the minimum and maximum reconstruction errors of u within duration $t = 1, 2, \dots, L^{(u)}$, respectively. The target health degradation factors trained from all the available samples are adopted to predict the values $\hat{\theta}$ and $\hat{\theta}_0$.

C. RUL Estimation With Degradation Factor

Let the notation H denote the health degradation value/plot. The health degradation factor for a single time step u^* is compared to all values throughout the lifetime $u \in U$. The prediction and training samples may be collected from different number of cycles to generate the same degradation factor. At any time step, the number of remaining cycles from the training samples after the last cycle of the prediction provides the RUL estimate for the test instance. Let u^* be a test instance and u a train instance. Specifically, we consider the following cases for degradation matching-based RUL estimation.

1) *Multiple Health Degradation Trends With High Similarity*: The health degradation plot $H^{(u^*)}$ may have high similarity with $H^{(u)}(t, L^{(u^*)})$ for multiple values of time-lag t . It is partially to only consider the RUL estimations for the time step t with a minimum Euclidean distance between the healthy degradations $H^{(u^*)}$ and $H^{(u)}(t, L^{(u^*)})$. Therefore, we conduct multiple RUL predictions for u^* based on the whole lifetime u . The multiple predicted RUL values for each time step are assigned with weight values proportional to the degradation similarity to achieve the final estimated value.

2) *Different Initial Health Status*: The initial healthy status of machines/components varies significantly because of various factors in the manufacturing process. The healthy similarity for a machine lifetime is divided by the mean value of the first-part health degradation values (e.g., the first quartile of total cycles). Furthermore, a time-lag t is tolerable so that the health degradation values of u^* may be close to the healthy degradation factor of $H^{(u)}(t, L^{(u^*)})$ at time t .

3) *Nonmonotonic Health Degradation Trend*: Health degradation plots sketched with linear regression are nonmonotonic because of the noisy sensory readings. The weighted moving average smoothing is used to minimize the estimation error in the health degradation factors.

4) *Maximum RUL Estimation Value*: It is challenging to estimate RUL if a machine is in good health status or at the beginning of the life cycle. Therefore, the maximum RUL value for each training sample is limited to R_{\max} . The maximum RUL value for the sample u^* -based similarity matching with sample u is limited by $L^{(u)} - L^{(u^*)}$. This indicates that the maximum RUL value for any training sample u is limited at the total length $\hat{R}^{(u^*)} + L^{(u^*)} \leq L_{\max}$, in which L_{\max} represents the maximum length of the training sample.

We define similarity health degradation trends u^* and training sample u at time step t as

$$s(u^*, u, t) = \exp\left[\frac{-d^2(u^*, u, t)}{\lambda}\right]$$

in which $-d^2(u^*, u, t)$ is calculated as follows:

$$d^2(u^*, u, t) = \frac{1}{L^{(u^*)}} \sum_{i=1}^{L^{(u^*)}} (h_i^{(u^*)} - h_{i+t}^{(u)})^2.$$

$d^2(u^*, u, t)$ represents the squared Euclidean distance between $H^{(u^*)}(1, L^{(u^*)})$ and $H^{(u)}(t, L^{(u^*)})$ subject to $\lambda > 0$, $t \in \{1, 2, \dots, \tau\}$, and $t + L^{(u^*)} \leq L^{(u)}$. In this formula, λ indicates the similarity level. A small value of λ means large difference in s . The predicted RUL value for u^* based on the healthy degradation for u and for time step t is calculated with $\hat{R}^{(u^*)}(u, t) = L^{(u)} - L^{(u^*)} - t$.

The predicted value $\hat{R}^{(u^*)}(u, t)$ is allocated with a weight of $s(u^*, u, t)$ in order to obtain weighted average estimate $\hat{R}^{(u^*)}$

$$\hat{R}^{(u^*)} = \frac{\sum s(u^*, u, t) \cdot \hat{R}^{(u^*)}(u, t)}{\sum s(u^*, u, t)} \quad (15)$$

in which the sum is for the combinations of u and t that satisfy the conditions of $s(u^*, u, t) \geq \alpha \cdot s_{\max}$ and $s_{\max} = \max_{u \in U, t \in \{1, \dots, \tau\}} \{s(u^*, u, t)\}$, $0 \leq \alpha \leq 1$.

IV. PERFORMANCE EVALUATION

This section presents the performance evaluation of the DELTA framework with the open-source FEMTO [14] data set. To validate the efficacy of the proposed solution, we also compare with the results of the mainstream RUL prediction algorithms, e.g., MLP, XGBoost, RFR, SVR, LSTM, and CNN. We have also included the performance comparison with RUL prediction algorithms especially developed for FEMTO bearing data set, e.g., the RNN [34], two-stage [36], and collaborative prediction [35]. Specifically, the implementations of CNN, LSTM, SVR, RNN [34], and two-stage [36] are explained as follows.

- 1) *SVR [4]*: The SVR uses the same RBF kernel as the SVM, with the objective to find the objective RUL as a continuous value (instead of a discrete value as the status classification). Note that this SVR algorithm also represents the solution without using multiple stages.
- 2) *CNN [5]*: The CNN is firstly adopted for the RUL prediction. This network architecture is composed of multivariate time-series input, two convolutional filtering layers, two pooling filtering layers, and one fully connected layer.
- 3) *LSTM [31]*: The LSTM predicts RUL values based on the previous estimations. Therefore, the first several predicted values are configured as max RUL values. Furthermore, we set the sliding window size as 50.
- 4) *RNN [34]*: In the RNN approach [34], six similarity-relevant features combined with eight classical time-frequency features in order to form an original feature set. Then, the most sensitive features are selected from the original feature set with monotonicity and correlation metrics. These selected features are fed into an RNN to construct the health indicator.
- 5) *Two-Stage [36]*: The two-stage data-driven approach [36] first calculates the deviation of multiple statistics of vibration signals of a bearing from a known healthy state. Then, a prediction stage based on an expectation-maximization algorithm and an enhanced Kalman filter are used to estimate the bearing RUL.

A. Evaluation Metrics and Experimental Data

We adopt two most widely used metrics for the performance evaluation in RUL prediction, i.e., score [30] and RMSE.

Score: This evaluation metric is defined by the authors of the C-MAPSS data set [30] and is widely used in existing studies [4], [5] for performance comparison. The mathematical expression is given as follows:

$$\text{Score} = \begin{cases} \sum_{i=1}^n \exp\left\{-\frac{d}{a_1}\right\} - 1, & \text{if } d < 0 \\ \sum_{i=1}^n \exp\left\{-\frac{d}{a_2}\right\} - 1, & \text{if } d \geq 0 \end{cases} \quad (16)$$

where n is the total number of units under test, $d = \text{Estimated RUL} - \text{True RUL}$, $a_1 = 10$, and $a_2 = 13$.

Root Mean Square Error (RMSE): RMSE is also a commonly used metric to compare the accuracy of competing

TABLE I
TRAINING AND TESTING DATA SETS OF FEMTO

Data-sets	FEMTO 1	FEMTO 2
Learning sets	Bearing 1-1	Bearing 2-1
	Bearing 1-2	Bearing 2-2
Testing sets	Bearing 1-3	Bearing 2-3
	Bearing 1-4	Bearing 2-4
	Bearing 1-5	Bearing 2-5
	Bearing 1-6	Bearing 2-6
	Bearing 1-7	Bearing 2-7
Load (N)	4000	4200
Speed (rpm)	1800	1650

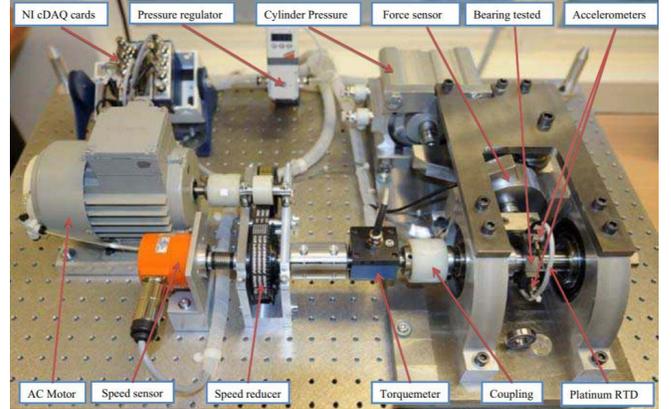


Fig. 3. Overview of the PRONOSTIA testbed.

algorithms. In particular, the mathematical expression of RMSE is given as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\text{Estimated RUL} - \text{True RUL})^2}{n}}$$

Table I lists the training and testing data sets of FEMTO. These data sets used in the bearing degradation experiments are collected in different operating conditions in terms of load and rotation speed. Specifically, the FEMTO data sets 1-1-1-7 are collected in the load of 4000 N and the rotation speed of 1800. The FEMTO data sets 2-1-2-7 are collected in the load of 4200 N and the rotation speed of 1650. As explained in Table I, the bearings 1-1 and 1-2 are adopted as the training/learning data sets. The five data sets from bearings 1-3-1-7 are used for testing/evaluation.

The algorithm performance comparison is conducted on an accelerated aging platform named PRONOSTIA, as presented in Fig. 3. This platform is specially developed for the simulation of the ball-bearings degradation process. The rotating part includes the asynchronous motor with a gearbox and its two shafts. The first shaft is near the motor, and the second shaft is placed at the ride side of the incremental encoder. The Bearing's operation conditions are determined by instantaneous measures of the radial force applied on the bearing, the rotation speed of the shaft handling the bearing and of the torque inflicted on the bearing. Each of these three analog measures is acquired at a frequency equal to 100 Hz. The characterization of the bearing's degradation is based on two data types of sensors: vibration and temperature. In the existing literature, there are different simulation methods

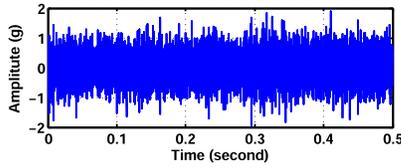


Fig. 4. Vibration signals for 0.5 s of the bearing 1 – 2.

TABLE II
COMPARISON OF RMSE ON FEMTO 1– DATA SETS

Algorithms \ Datasets	1-3	1-4	1-5	1-6	1-7
LR	307.8	860.3	678.3	1023.7	1685.7
SVR	258.6	1102.6	885.7	1075.3	1127.3
XGBoost	113.5	116.7	476.2	226.7	343.6
CNN	151.3	197.3	637.5	527.8	419.6
LSTM	108.7	145.3	512.6	363.7	387.5
RNN	105.2	138.6	523.8	382.4	397.3
two-stage	176.2	218.3	558.3	435.6	475.6
DELTA	80.9	112.3	127.5	113.6	128.6

for the degradation process of ball bearings. This machine degradation can be simulated by introducing impurities into the lubricant or scratching the bearing surface.

Seventeen bearings are tested under three different operation conditions, in which two bearings under each operation condition are used for training, and the others are used for testing. The operation conditions in terms of load and speed are shown in Table I. The accelerometers are fixed on the outer race of the bearings, and vibration signals are captured. The sampling frequency is 25.6 kHz. Each sample includes 2560 points (i.e., 0.1 s), and sampling is repeated every 10 s. Depending on the uncertainty of the degradation processes and the diversity of the different bearings, the faulty patterns of each bearing are different from each other. In order to avoid the propagation of damages to the whole experimental system, tests are stopped when the amplitude of the vibration signal exceeds 20 g. Correspondingly, the time when the amplitude of the vibration signal exceeds 20 g is determined as the end-of-life. Fig. 4 shows the vibration signals of bearing 1 – 5 in [0, 0.5] s. It is obvious that the vibration signals include two distinctly different stages, i.e., the normal operation stage and the degradation stage. The vibration signals keep stable during the normal operation stage and increase rapidly during the degradation stage. Taking Bearing 1 for example, two segments of signals in different stages are amplified to present more detailed information. It is observed that there exist periodic impact components in the degradation stage.

B. Evaluation Results

We use the piecewise linear (PWL) RUL [5], [33] instead of actual RUL in the training and prediction processes. PWL RUL means the actual RUL will be set to the max RUL if it is larger than this max value. The max RUL is configured as 16000 in all the experiments.

The RMSE and score values of all the evaluated algorithms for data sets 1 – 3–1 – 7 are presented in Tables II and III. As the complexity and amount of the sensory data increase,

TABLE III
COMPARISON OF SCORE ON FEMTO 1– DATA SETS

Algorithms \ Datasets	1-3	1-4	1-5	1-6	1-7
LR	2365.7	837.2	1785.3	1876.8	2874.3
SVR	1607.2	1693.5	1873.2	2017.3	3417.2
XGBoost	927.5	927.6	1235.7	1137.6	2576.6
CNN	953.6	1143.7	1728.8	1937.8	2617.3
LSTM	908.6	1082.9	1368.7	1673.8	2793.8
RNN	932.8	1065.7	1293.5	1732.9	2678.5
two-stage	927.5	927.6	1463.8	1258.3	2682.9
DELTA	508.7	723.2	973.6	669.8	1817.4

TABLE IV
COMPARISON OF RMSE ON FEMTO 2– DATA SETS

Algorithms \ Datasets	2-3	2-4	2-5	2-6	2-7
LR	478.6	893.5	682.7	1047.8	918.4
SVR	1243.7	1187.3	895.7	1053.8	1143.2
XGBoost	278.2	119.7	518.4	236.8	359.6
CNN	513.2	289.3	543.7	578.9	573.8
LSTM	587.2	247.8	583.7	513.7	592.6
RNN	489.3	253.6	613.5	549.3	587.2
two-stage	632.7	337.2	573.5	592.6	642.8
DELTA	131.9	86.7	143.7	128.7	253.9

TABLE V
COMPARISON OF SCORE ON FEMTO 2– DATA SETS

Algorithms \ Datasets	2-3	2-4	2-5	2-6	2-7
LR	1217.3	1753.2	2017.3	2865.8	2967.3
SVR	1965.2	1483.5	1573.8	2967.3	2618.4
XGBoost	1493.2	1875.6	1143.7	1407.8	1328.9
CNN	1567.2	18719.3	1327.4	1675.3	1817.9
LSTM	1689.2	1758.3	1517.3	1482.7	1387.2
RNN	1582.3	1863.5	1582.8	1569.3	1453.6
two-stage	1632.5	1982.6	1258.2	1382.6	1317.3
DELTA	819.3	752.8	887.6	397.5	987.2

the scores of all the evaluated algorithms become higher (i.e., the overall accuracy decreases). For the data set 1 – 4, XGBoost achieves approximately the same accuracy level as the proposed DELTA, but the performance gaps become larger on the other three data sets.

Tables IV and V present the RMSE and score values obtained on the data sets 2 – 3–2 – 7. Different from the results in the above FEMTO 1 data sets, the accuracy of SVR is generally lower than the other competing algorithms. Overall, our proposed DELTA performs better than all the other competing schemes in the evaluated data sets. XGBoost and LSTM achieve higher accuracy than the remaining three models of CNN, SVR, and LR.

The microscopic RUL prediction for two data sets 1 – 3 and 1 – 5 is demonstrated in Fig. 5. The objective is to compare the differences in the instantaneous RUL values with two typical reference models of XGBoost and LSTM. As the results indicate, the proposed DELTA predicts RUL values that are close to the ground-truth plot with lower variations than the competing algorithms. The mean and smoothed values of the RUL prediction values are also presented in Fig. 5(a)–(c) and (e)–(g) for comparison. Fig. 5(d) and (h) shows the composite plots of smoothed values from the three approaches.

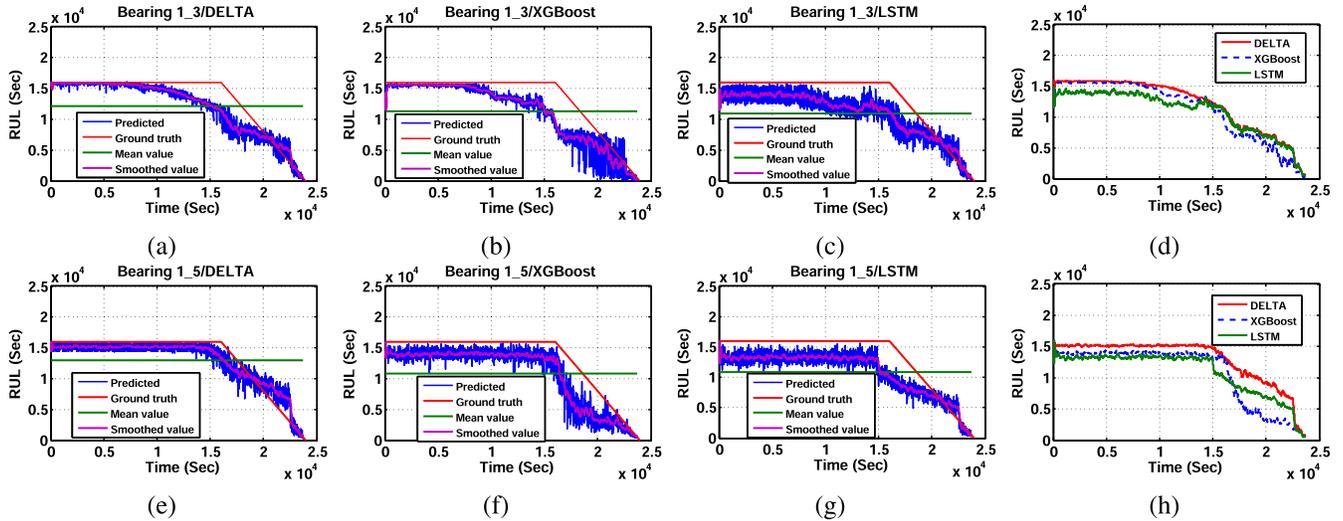


Fig. 5. Microscopic RUL prediction results of the FEMTO data set.

TABLE VI
SCORE AND RMSE OF THE EVALUATED ALGORITHMS

Score Value				
Algorithms \ Dataset	1-3	1-7	2-3	2-7
DELTA (w/o classification)	913.5	1983.4	1026.8	1265.3
DELTA (w/o AE)	735.8	1967.3	887.3	1127.3
DELTA	508.7	1817.4	819.3	987.2
RMSE Value				
Algorithms \ Dataset	1-3	1-7	2-3	2-7
DELTA (w/o classification)	113.8	145.6	167.4	327.5
DELTA (w/o AE)	98.6	122.8	147.5	297.3
DELTA	80.9	128.6	131.9	283.9

C. Model Ablation Study

In this section, we evaluate the performance contributions of the healthy stage classification and AE components to DELTA. These two modules are the two most important components in our proposed DELTA. Specifically, we first disable the stage classification component and compare the accuracy with the baseline DELTA using the FEMTO data set. The evaluation result is presented in Table VI. The baseline DELTA algorithm with all the components is expressed as DELTA, while the scheme without the stage classification component is denoted with DELTA (w/o classification). The difference in the results indicates the importance of the healthy stage classification.

As explained previously, we adopt the AE for the health degradation factor calculation. This is because AE exploits the latent representations/variables in the available sensory data to predict the degradation. The method of DELTA w/o AE means using only the LSTM (replacing the LSTM AE) for the health degradation factor calculation. As the results indicate, it is important to adopt AE in our proposed DELTA framework since there is the substantial difference if excluding this component.

The results in Table VI indicate that the classification module provides more performance contributions than the

AE in our proposed DELTA. This is because the DELTA (w/o classification) achieves the lowest accuracy in the three models.

V. CONCLUSION AND DISCUSSION

With the technological advancements in industrial IoT, RUL prediction plays a pivotal role in predictive system maintenance. This article develops a joint classification-regression method dubbed DELTA to adaptively predict RUL in multiple stages. The objective is to accurately diagnose the faults and reduce the cost of predictive maintenance. This article proposes a DELTA scheme that predicts RUL in multiple states to improve the estimation accuracy. The approach was evaluated using turbofan engines and shaft production system data. Results demonstrate appreciable performance improvements over the competing algorithms. Furthermore, DELTA is a flexible framework with substitutable components in the key modules of healthy stage identification and degradation factor estimation.

As future work, we will consider adopting the federated learning framework for distributed RUL prediction to address the data privacy and transmission issues.

REFERENCES

- [1] D. Wang, F. Yang, K.-L. Tsui, Q. Zhou, and S. J. Bae, "Remaining useful life prediction of lithium-ion batteries based on spherical cubature particle filter," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 6, pp. 1282–1291, Jun. 2016.
- [2] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, "Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2306–2318, Oct. 2017.
- [3] R. K. Singleton, E. G. Strangas, and S. Aviyente, "Extended Kalman filtering for remaining-useful-life estimation of bearings," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1781–1790, Mar. 2015.
- [4] R. Khelif, B. Chebel-Morello, S. Malinowski, E. Laajili, F. Fnaiech, and N. Zerhouni, "Direct remaining useful life estimation based on support vector regression," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2276–2285, Mar. 2017.
- [5] G. S. Babu, P. Zhao, and X. L. Li, "Deep convolutional neural network based regression approach for estimation of remaining useful life," in *Proc. Int. Conf. Database Syst. Adv. Appl.* Cham, Switzerland: Springer, Apr. 2016, pp. 214–228.

- [6] 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.
- [7] N. Gebrael, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: A neural network approach," *IEEE Trans. Ind. Electron.*, vol. 51, no. 3, pp. 694–700, Jun. 2004.
- [8] A. Soualhi, H. Razik, G. Clerc, and D. D. Doan, "Prognosis of bearing failures using hidden Markov models and the adaptive neuro-fuzzy inference system," *IEEE Trans. Ind. Electron.*, vol. 61, no. 6, pp. 2864–2874, Jun. 2014.
- [9] H.-E. Kim, A. C. C. Tan, J. Mathew, and B.-K. Choi, "Bearing fault prognosis based on health state probability estimation," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5200–5213, Apr. 2012.
- [10] S. S. H. Zaidi, S. Aviyente, M. Salman, K.-K. Shin, and E. G. Strangas, "Prognosis of gear failures in DC starter motors using hidden Markov models," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1695–1706, May 2011.
- [11] W. Wu, J. Hu, and J. Zhang, "Prognostics of machine health condition using an improved ARIMA-based prediction method," in *Proc. 2nd IEEE Conf. Ind. Electron. Appl.*, May 2007, pp. 1062–1067.
- [12] J. B. Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, and F. Fnaiech, "Accurate bearing remaining useful life prediction based on weibull distribution and artificial neural network," *Mech. Syst. Signal Process.*, vols. 56–57, pp. 150–172, May 2015.
- [13] M. Feldman, "Hilbert transform in vibration analysis," *Mech. Syst. Signal Process.*, vol. 25, no. 3, pp. 735–802, 2011.
- [14] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (PHM)*, Jun. 2012, pp. 1–8.
- [15] A. Mosallam, K. Medjaher, and N. Zerhouni, "Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction," *J. Intell. Manuf.*, vol. 27, no. 5, pp. 1037–1048, Oct. 2016.
- [16] M. E. Orchard, P. Hevia-Koch, B. Zhang, and L. Tang, "Risk measures for particle-filtering-based state-of-charge prognosis in lithium-ion batteries," *IEEE Trans. Ind. Electron.*, vol. 60, no. 11, pp. 5260–5269, Nov. 2013.
- [17] Q. Miao, L. Xie, H. Cui, W. Liang, and M. Pecht, "Remaining useful life prediction of lithium-ion battery with unscented particle filter technique," *Microelectron. Rel.*, vol. 53, no. 6, pp. 805–810, Jun. 2013.
- [18] E. Ramasso, M. Rombaut, and N. Zerhouni, "Joint prediction of continuous and discrete states in time-series based on belief functions," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 37–50, Feb. 2013.
- [19] K. Javed, R. Gouriveau, and N. Zerhouni, "A new multivariate approach for prognostics based on extreme learning machine and fuzzy clustering," *IEEE Trans. Cybern.*, vol. 45, no. 12, pp. 2626–2639, Dec. 2015.
- [20] F. Xue, P. Bonissone, A. Varma, W. Yan, N. Eklund, and K. Goebel, "An instance-based method for remaining useful life estimation for aircraft engines," *J. Failure Anal. Prevention*, vol. 8, no. 2, pp. 199–206, Apr. 2008.
- [21] R. Khelif, S. Malinowski, B. C. Morello, and N. Zerhouni, "Unsupervised kernel regression modeling approach for RUL prediction," in *Proc. Eur. Conf. Prognostics Health Manage. Soc.*, Jul. 2014, pp. 1–7.
- [22] R. Khelif, S. Malinowski, B. Chebel-Morello, and N. Zerhouni, "RUL prediction based on a new similarity-instance based approach," in *Proc. IEEE Int. Symp. Ind. Electron.*, Jun. 2014, pp. 2463–2468.
- [23] E. Ramasso, "Investigating computational geometry for failure prognostics," *Int. J. Prognostics Health Manage.*, vol. 5, pp. 1–18, Nov. 2014.
- [24] T. Qin, S. Zeng, and J. Guo, "Robust prognostics for state of health estimation of lithium-ion batteries based on an improved PSO-SVR model," *Microelectron. Rel.*, vol. 55, nos. 9–10, pp. 1280–1284, Sep. 2015.
- [25] 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.
- [26] M. A. Patil *et al.*, "A novel multistage support vector machine based approach for lithium-ion battery remaining useful life estimation," *Appl. Energy*, vol. 159, pp. 285–297, Dec. 2015.
- [27] T. H. Loutas, D. Roulias, and G. Georgoulas, "Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic E-support vectors regression," *IEEE Trans. Rel.*, vol. 62, no. 4, pp. 821–832, Dec. 2013.
- [28] H. Dong, X. Jin, Y. Lou, and C. Wang, "Lithium-ion battery state of health monitoring and remaining useful life prediction based on support vector regression-particle filter," *Power Sources*, vol. 271, pp. 114–123, Dec. 2014.
- [29] Y. Lei, N. Li, and J. Lin, "A new method based on stochastic process models for machine remaining useful life prediction," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 12, pp. 2671–2684, Dec. 2016.
- [30] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Proc. Int. Conf. Prognostics Health Manage. (PHM)*, Oct. 2008, pp. 1–9.
- [31] H. Miao, B. Li, C. Sun, and J. Liu, "Joint learning of degradation assessment and RUL prediction for aeroengines via dual-task deep LSTM networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5023–5032, Sep. 2019.
- [32] M. Xia, T. Li, T. Shu, J. Wan, C. W. de Silva, and Z. Wang, "A two-stage approach for the remaining useful life prediction of bearings using deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 6, pp. 3703–3711, Jun. 2019.
- [33] Z. Chen, M. Wu, R. Zhao, F. Guretno, R. Yan, and X. Li, "Machine remaining useful life prediction via an attention based deep learning approach," *IEEE Trans. Ind. Electron.*, vol. 68, no. 3, pp. 2521–2531, Mar. 2021, doi: [10.1109/TIE.2020.2972443](https://doi.org/10.1109/TIE.2020.2972443).
- [34] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, May 2017.
- [35] L. Ren, J. Cui, Y. Sun, and X. Cheng, "Multi-bearing remaining useful life collaborative prediction: A deep learning approach," *J. Manuf. Syst.*, vol. 43, pp. 248–256, Apr. 2017.
- [36] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K.-L. Tsui, "A two-stage data-driven-based prognostic approach for bearing degradation problem," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 924–932, Jun. 2016.



Ji-Yan Wu received the Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2014.

His current research is on sensory data analytics, machine prognostic techniques, and federated learning.



Min Wu (Senior Member, IEEE) received the B.S. degree in computer science from the University of Science and Technology of China, Hefei, China, in 2006, and the Ph.D. degree in computer science from Nanyang Technological University, Singapore, in 2011.

He is currently a Senior Scientist with Data Analytics Department, Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. His current research interests include machine learning, data mining, and bioinformatics.

Dr. Wu was the recipient of the Best Paper Awards in the International Conference on Bioinformatics in 2016 and International Conference on Database Systems for Advanced Applications in 2015. He also won the International Joint Conferences on Artificial Intelligence Competition on repeated buyers prediction in 2015.



Zhenghua Chen received the B.Eng. degree in mechatronics engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2011, and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2017.

He is currently a Scientist with the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. His research interests include data analytics in smart buildings, ubiquitous computing, Internet of Things, machine learning, and deep learning.



Xiao-Li Li (Senior Member, IEEE) received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China.

He is currently a Principal Scientist with the Institute for Infocomm Research, A*STAR, Singapore. He is also an Adjunct Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He has authored and coauthored more than 180 high quality articles. His research interests include data

mining, machine learning, artificial intelligence, and bioinformatics.

Prof. Li has been serving as a (Senior) Program Committee Member/Workshop Chair/Session Chair in leading data mining and artificial-intelligence-related conferences, including KDD, ICDM, SDM, PKDD/ECML, WWW, IJCAI, AAAI, ACL, and CIKM. He has won numerous best article/benchmark competition awards.



Ruqiang Yan received the M.S. degree in precision instrument and machinery from the University of Science and Technology of China, Hefei, China, in 2002, and the Ph.D. degree in mechanical engineering from the University of Massachusetts, Amherst, MA, USA, in 2007.

From 2009 to 2018, he was a Professor with the School of Instrument Science and Engineering, Southeast University, Nanjing, China. He joined the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China, in 2018. He has authored

and coauthored two books and more than 200 articles in technical journals and conference proceedings. He holds 28 patents. His research interests include data analytics, machine learning, and energy-efficient sensing and sensor networks for the condition monitoring and health diagnosis of large-scale, complex, dynamical systems.

Dr. Yan became a fellow of the American Society of Mechanical Engineers, in 2019. He was the recipient of several honors and awards including the IEEE Instrumentation and Measurement Society Technical Award in 2019, the New Century Excellent Talents in University Award from the Ministry of Education in China, in 2009, and multiple best paper awards. He is an Associate Editor-in-Chief for the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT and an Associate Editor for the IEEE SYSTEMS JOURNAL.