A Deep Learning Feature Fusion Based Health Index Construction Method for Prognostics Using Multiobjective Optimization

Zhen Chen, *Member, IEEE*, Di Zhou, Enrico Zio, *Senior Member, IEEE*, Tangbin Xia, *Member, IEEE*, and Ershun Pan

***Abstract*—Degradation modeling and prognostics serve as the basis for system health management. Recently, various sensors provide plentiful monitoring data that can reflect the system status. A multitude of feature fusion techniques based on multisensor data have been proposed to generate a composite health index (HI) for prognostics, which can represent the underlying degradation mechanism. Most existing methods have used linear fusion models and neglected the practical requirements for HI construction, which are insufficient to reveal the nonlinear relations among features and difficult to obtain accurate HIs for complicated systems. This study proposes a novel feature fusion-based HI construction method with deep learning and multiobjective optimization. Multiple degradation features are fused by a deep neutral network (DNN). Several desired properties that the HIs should have for prognostics are adopted to formulate the objective functions of DNN training. To balance the spatial complexity and performance of the fusion model, a multiobjective optimization model is generated for training the DNN. Then, a generalized nonlinear Wiener process model is used to predict the remaining useful life with the resulted HIs. Finally, two cases are analyzed to illustrate the effectiveness and robustness of the proposed method.**

***Index Terms*—Deep learning, feature fusion, health index (HI), multiobjective optimization, remaining useful life (RUL).**

NOTATIONS

h_{ij}	Constructed HI of unit i at time t_j.
Mon(·)	Monotonicity.
Tre(·)	Trendability.
Rob(·)	Robustness.
Sca(·)	Scale similarity.
Con(·)	Consistency.
$x_{q,j}$	The qth degradation feature at time t_j.
$\boldsymbol{w}$	Vector of weights of DNN.
$\boldsymbol{b}$	Vector of bias of DNN.
L	Number of the hidden layers of DNN.
d_l	Number of neurons at layer l of DNN.
$J_1(\boldsymbol{w}, \boldsymbol{b})$	Objective function for representation performance.
$J_2(L, \boldsymbol{d})$	Objective function for spatial complexity.
A(·)	Cost-effectiveness ratio function.
$\mu(t; \theta)$	Drift function.
T	First passage time.
R_j	The remaining useful life at time t_j.

Manuscript received 24 April 2022; revised 7 July 2022; accepted 11 October 2022. The work was supported in part by the National Natural Science Foundation of China under Grant 72001138, Grant 52005327, and Grant 72071127 and in part by the National Key Research and Development Program of China under Grant 2020YFB1711100. Associate Editor: Shubin Si. *(Corresponding author: Ershun Pan.)*

Zhen Chen, Tangbin Xia, and Ershun Pan are with the Department of Industrial Engineering and Management and with the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: chenzhendr@sjtu.edu.cn; xtbxtb@sjtu.edu.cn; pes@sjtu.edu.cn).

Di Zhou is with the College of Mechanical Engineering, Donghua University, Shanghai 200051, China (e-mail: zhoudi@sjtu.edu.cn).

Enrico Zio is with the Centre for Research on Risk and Crises, Ecole des Mines, ParisTech, PSL University, 75006 Paris, France, and also with the Department of Energy, Politecnico di Milano, 20133 Milan, Italy (e-mail: enrico.zio@polimi.it).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TR.2022.3215757.

Digital Object Identifier 10.1109/TR.2022.3215757

I. INTRODUCTION

REMAINING useful life (RUL) prediction has a vital role in prognostics and health management, which becomes an effective technique to reveal the health status of industrial systems and reduce the losses of production and economy [1], [2]. Since the status of many systems, such as machinery and electronic devices, degrades gradually, the RUL is commonly defined as the residual time when the irreversible accumulation of damage reaches a critical threshold from the current state.

With the rapid development of sensing techniques, a great deal of sensor data can be collected from the condition monitoring system. The sensing data related to degradation processes is known as the degradation signals [3]. These signals can reveal the system health states directly or indirectly. Many studies modeled the degradation processes with a single raw signal, such as crack length, vibration amplitude, and battery capacity [4], [5]. If the failure mechanism is clear, a single direct signal may be capable of revealing the underlying degradation characteristics of the system status. However, in many industrial systems with complicated structure and under changeable operating environment, a single raw degradation signal cannot completely contain sufficient information for degradation modeling and RUL prediction.

Accordingly, multiple monitoring signals have been commonly used to extract degradation features [6], [7]. Based on the extracted multiple features, there exist two types of methods for the RUL prediction. The first type, known as end-to-end manner, is directly modeling the relationships between the RUL and

multiple degradation features [8]. In recent years, deep learning technology grow rapidly and exhibit its excellent performance in the RUL prediction [9], [10], [11]. Liu et al. [12] developed a feature-attention-based end-to-end approach to predict the RUL. Ma and Mao [13] stacked multiple convolution-based long short-term memory network and formed an encoding-forecasting architecture for RUL prediction. Jin et al. [14] used a bidirectional long short-term memory (LSTM)-based two-stream network for RUL prediction. All these methods are supervised methods, where multiple features were taken as model inputs and the occurred lifetime or RUL percentages were directly used as model outputs, i.e., labels. This means that the predicted RULs would linearly decreases to zero and be used to represent the system status. However, the real degradation processes in practice are commonly nonlinear and complex. Additionally, as the progression of the underlying status is also not provided, it is challenging for engineers to rely on the predicted RUL for making further decisions (e.g., spare part logistics, inventory control, and maintenance).

The methods of the second type employ feature fusion techniques to construct a composite health index (HI) and then predict the RULs by degradation models [15]. As the one-dimensional (1-D) HI provides a straightforward visualization and intuitive interpretation of the underlying degradation status of a system [16], this type of methods is promising to better capture the evolution mechanism of degradation and improve the decision-making in practice. Then, the RUL prediction can be facilitated by using various existing univariate prognostics. Furthermore, compared with the second type of methods, it is difficult for the first type of methods to perform well in the early stage of the system operation, when degradation phenomenon is always not noticeable. Although the degradation features are almost unchanged in the early stage, the predicted RUL reduces linearly. That is, the corresponding starting point of rapid degradation stage is hard to be identified and the first type of methods has no ability to predict the RUL at the early stage [17].

Many previous studies have considered feature fusion techniques for the HI construction [18], [19], [20], [21]. However, most of these studies ignored the actual requirements and desired properties of degradation modeling and prognostics in the HI construction. To achieve more accurate RUL predictions by simpler prognostics, some researchers have indicated some desired properties, which the constructed HI is expected to have [22]. Khanh and Kamal [23] summarized five intrinsic natures of HI and three extern correlations for the HI construction, such as monotonicity, trendability, robustness, and mutual information. Several evaluation criteria were also presented to measure the performance of the prognostic results. Some studies only considered a part of the abovementioned essential properties in the HI construction step. Liao [24] used genetic programming which formulates the fitness function with monotonicity to generate the HI for bearings. Liu et al. [15] developed a composite HI based on monotonicity and variance by using a linear weighted sum of multiple features. Furthermore, Liu et al. [25] defined a consistency property to optimize the constructed HIs. Yan et al. [26] proposed a linear programming-based feature fusion approach to generate the HI with the monotonicity, trendability, and robustness properties. Wang et al. [27] proposed a deep learning-based data fusion framework with the properties of monotonicity and range information. Furthermore, involving more essential properties may yield more efficient HI construction results. To achieve an HI that represents the degradation processes more accurately, Nguyen and Medjaher [23] used a simple combination of all eight properties as one fitness function of genetic programming for feature fusion. The results also showed that the HIs based on multiple essential properties performed best in prognostics than those based on simpler construction criteria. Additionally, more scholars preferred using these properties in feature selection or HI evaluation [28], instead of discovering the most appropriate HIs by optimization [29]. This cannot ensure the resulted HIs to characterize the failure progression effectively.

Besides the consideration of multiple desired properties of HI, how to select an appropriate fusion function is another key issue. The weighted linear combination of multiple features is most used fusion function [15], [26]. Chen et al. [30] adopted a nonlinear function to fuse the weighted sum of multiple features for formulating a composite HI. Since the failure mechanisms of systems are complex and the relationships among multiple features are not clearly revealed, it is difficult for those common functions (linear weighted function, exponential function, sigmoid function, etc.) to accurately fuse the features and further represent the underlying degradation status. Due to the great nonlinear approximation ability of deep learning algorithms [31], Wang et al. [27] established a deep neural network (DNN) to fuse multiple sensor data for degradation modeling. Li et al. [32] proposed a shape-constrained neural network for data fusion and HI construction, where the loss function was formulated based on both the monotonicity and the failure time variability. Wang et al. [17] used a generic indirect deep learning algorithm for the HI construction with multiple features based on the concept of monotonicity and failure threshold. However, only two desired properties were considered in those previous studies and the topological parameters of deep networks were determined by cross validation, which significantly affected the application accuracy and efficiency. By far the deep learning-based fusion models have been not researched thoroughly.

To address all abovementioned concerns, this study develops a novel HI construction framework with DNN and multiobjective optimization. Several desired properties that an HI is expected to have in prognostic tasks are used to formulate the function of representation performance. Considering that the representation performance of HI and the spatial complexity of DNN simultaneously, a multiobjective optimization model (MOM) is developed to train the DNN model. To verify the predictability of the constructed HIs, a generalized nonlinear Wiener process (WP) model based on kernel functions is built.

Note that, since the actual health status is often unavailable and cannot be directly measured from sensor data, the application range and accuracy of most supervised end-to-end methods are limited, which often use manually labeled values instead of actual RULs as the response variable. Obviously, the proposed method is apparently distinct from the supervised algorithm. On

the other side, the proposed method also differs from the unsupervised task. The reason is that some domain knowledge, which reveals the underlying degradation mechanism is introduced into the HI construction, such as monotonicity and failure threshold. However, exiting deep learning-based end-to-end methods are black-box models regardless of domain knowledge. Song and Liu [33] proposed an indirect supervised learning method that although the underlying degradation mechanism was unavailable, the failure time of historical units, domain knowledge and monitoring data were used to determine the unobservable labels indirectly.

The main contributions of this article are as follows.

1) A DNN-based feature fusion model is developed to construct composite HIs based on multiple degradation features, which can represent nonlinear characteristics and complex relationships among different feature sequences effectively.
2) The MOM is formulated to estimate the topological parameters and unknown coefficients of DNN, which considers the representation performance of HI and the spatial complexity of DNN simultaneously.
3) To discover the most suitable HIs, the domain knowledge including five key properties are used to formulate the objective of the MOM rather than evaluation metrics. A hierarchical self-adaptive differential evolution (HSADE) algorithm is also developed.
4) A generalized nonlinear WP model with kernel functions, that has higher generality and flexibility in nonlinear modeling without prior information, is proposed to predict the RUL distributions with the constructed HIs.

The rest of this article is organized as follows. Some desired properties of HIs for prognostics are introduced in Section II. Section III proposes a methodology of HI construction and prognostics. Section IV presents two illustrative examples. Finally, Section V concludes this article.

II. Desired Properties for HI

Some desired prosperities of HI have been presented in previous studies to construct an appropriate HI for prognostics. In addition, some metrics for assessing the HI performance have also been widely studied. This section aims to summarize the HI intrinsic prosperities and the evaluation metrics, which can be used for the HI construction and prognostic applications.

A. Desired Properties

When constructing an HI for describing the degradation process, the time-dependence characteristic and degradation evolution mechanism, and the accuracy and robustness goals of prognostics need to be considered. Motivated by these requirements, the intrinsic properties that desired for HI construction and their corresponding computational formulas are summarized as follows.

Property 1. Monotonicity: To capture a system's degradation mechanism more accurately and clearly, the trend of the HI should be monotonically increasing and decreasing.

Property 2. Trendability: As the system gradually degrades over time, an appropriate HI should be associated with the operation time. This property is called as trendability.

Property 3. Robustness: Due to sensor noises and variations in operating conditions, the degradation processes inevitably fluctuate. The robustness is employed to measure the stability or uncertainty of the HI.

Due to the randomness and noise, the monitoring signals are usually nonmonotonic. Thus, the constructed HI is expected to present an obvious trend for easily predicting the system's RUL. A higher trendability means the failure propagation over time can be charactered well by the HI, and thereby the HI has better prognosticability. Let $\boldsymbol{h}_{1:m_i}^{(i)} = [h_{i,1}, \ldots, h_{i,m_i}]^{\mathrm{T}}$ denote the ith HI trajectory, where $h_{i,j}$ is the HI at time t_j for sample i, $i = 1, \ldots, n$, $j = 1, \ldots, m_i$. Let $\mathrm{Mon}(\boldsymbol{h})$, $\mathrm{Tre}(\boldsymbol{h})$, and $\mathrm{Rob}(\boldsymbol{h})$ denote the monotonicity, trendability, and robustness, respectively, whose range values are all [0, 1]. Their calculation formulas can be found in [30]. $\mathrm{Mon}(\boldsymbol{h}) = 1$ means that all HI trajectories are strictly monotonic, while $\mathrm{Tre}(\boldsymbol{h}) = 1$ indicates that the HI trajectories are totally relevant to the system lifetime. The robustness score is expected to be close to 1 for lower fluctuating trajectories.

Property 4: Scale similarity. To enhance the generalization of constructed HIs, the range scales of all trajectories should be closer. The formulation of scale similarity is

$$\mathbf{Sca}\left(\boldsymbol{h}\right) = 1 - \frac{1}{n}\sum_{i=1}^{n}\frac{h_{\max} - \max\left(h^{(\mathrm{i})}\right) - h_{\min} + \min\left(h^{(\mathrm{i})}\right)}{h_{\max} - h_{\min} + \max\left(h^{(\mathrm{i})}\right) - \min\left(h^{(\mathrm{i})}\right)} \tag{1}$$

where $h_{\max}$ and $h_{\min}$ are the maximal and minimal values of all HI trajectories, respectively. $\mathbf{Sca}(\boldsymbol{h}) \in [0, 1]$.

Property 5: Consistency. Considering different HIs under the same operational condition, their failure thresholds should keep consist. The formulation of consistency is given by

$$\mathbf{Con}\left(\boldsymbol{h}\right) = \frac{n\exp\left(-\sigma\left(h_m\right)\right)}{\sum_i^n \left|h_{i,m_i} - h_{i,0}\right|} \tag{2}$$

where $\sigma(h_m)$ is the standard deviation (SD) of the terminal points of all HI trajectories, and $h_{i,0}$ is the initial value of ith trajectory. If all trajectories have the same initial and the same terminal points, the consistency reach its maximum, i.e., $\mathrm{Con}(\boldsymbol{h}) = 1$.

Even if all the units are under the same operational condition or have the same load profile, they may have different initial health states and degrade with different rates and variations due to unit-to-unit variability. This kind of uncertainties can be caused by the randomness of operation environment, and the inherent variations in raw materials or manufacturing processes. It is obvious that the heterogeneity can affect the modeling and prediction accuracy. Besides these five properties, the failure time of historical units is also considered in the HI construction. The first passage time (FPT) of the degradation trajectory exceeding a prefixed critical threshold is generally regarded as the failure time. The threshold is determined according to engineering practice and expert knowledge. For example, a lithium-ion battery is defined to fail when its capacity decreases to 70%–80% of the initial capacity. Thereby, the failure time

of identical units are probably different, which lead to another source of uncertainty. This type of uncertainty could affect the stability of prediction. To reduce the effects of these uncertainties on the prognostics, the properties 4 and 5 should be considered in the HI construction.

The scale similarity is used to represent the unit-to-unit heterogeneity. If the maximal and minimal values of different HI trajectories are the same, Sca($\boldsymbol{h}$) is equal to 1. It suggests that when the degradation processes of different units are projected to the constructed HI space, the change amplitudes of degradation are the same. Then, when we construct the HIs for testing units, their degradation trajectories will be fall into the same interval. That is, there is no difference in the severities of degradation for different units under the same operational condition and the heterogeneity can be decreased. This constraint can markedly enhance the generalization of the proposed HI construction method.

On the other side, the consistency represents the uncertainty resulted from the failure time of different units. In practice, the failure thresholds of identical systems are commonly the same. Moreover, some anomaly detection techniques can be used to identify the initial degradation points (IDPs) of historical units. Then, we can truncate the data of historical units to ensure only the observations from the initial degradation to failure are selected as training data. To ensure the HI values at the initial timepoint and at the final timepoint of all training units to respectively be 0 and 1, the data normalization need to be conducted to the constructed HIs for final output. Thus, the IDPs of the testing units can be detected which can make the initial HI equal to 0, and the testing units are expected to have the same failure threshold that corresponds to the HI value equal to 1. This makes Con($\boldsymbol{h}$) tending to 1. By using the data truncation and normalization, the consistency property is realized and the uncertainties will be sharply reduced.

Furthermore, to achieve the maximums of these properties, the optimization is essential. In most previous studies, a part of or all those above intrinsic prosperities are usually used to assess the HI performance after the construction phase, rather than incorporated into the HI construction step. Hence, the obtained HIs may be not optimal and accurate. To make up for this shortcoming, we will take these five properties as the objectives in the construction and optimization of HIs.

B. Performance Evaluation Metrics

Once the HIs are constructed, they will be substituted into the degradation models for predicting the RULs. Then, the most concern is the prognostic accuracy. There are several evaluation metrics that have been discussed in numerous studies.

Mean absolute relative error (MARE): Measuring the accuracy of the RUL prediction. Its calculation formula is

$$\text{MARE} = \frac{1}{m}\sum_{j=1}^{m}\frac{\left|R_j - R_j^*\right|}{R_j^*} \tag{3}$$

where R_j and R_j^* denote the predicted and the actual RUL values at time t_j, respectively.

Root mean square error (RMSE): Evaluating the mean root squared difference between the predicted and the actual RUL values, i.e.,

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(R_j - R_j^*\right)^2}. \tag{4}$$

Mean absolute deviation (MAD): Measuring the variability of the differences between the predicted RULs and their average values, i.e.,

$$\text{MAD} = \sqrt{\frac{1}{m-1}\sum_{j=1}^{m}\left(\Delta R_j - \Delta\bar{R}\right)^2} \tag{5}$$

where $\Delta R_j = R_j - R_j^*$ and $\Delta\bar{R}$ are the mean value of ΔR_j. The MAD can reflect the robustness of the predictions.

Prognostic score (PS): Since later predictions could result in serious system malfunctions in practical applications, the later predictions should be penalized more than the earlier predictions. Based on this preference, the formula of the PS metric is defined by an exponent function as follows

$$\text{PS} = \frac{1}{m}\sum_{j=1}^{m}\gamma\left(\Delta R_j\right) \tag{6}$$

where $\gamma(z) = \exp\{|z|/a_2\} - 1$, if $z \geq 0$; $\gamma(z) = \exp\{|z|/a_1\} - 1$, if $z < 0$, $a_1 > a_2 > 0$.

All abovementioned metrics tend to be zero when the prognostics based on the constructed HIs are perfect. As these performance metrics rely more on the validity of the degradation models, they are only considered in the evaluation step rather than the HI construction step in the following.

III. Methodology

A DNN-based feature fusion method is proposed here to construct a composite HI with multiple features. A multiobjective optimization problem based on the desired prosperities and the model spatial complexity is formulated to optimize the constructed HIs. To further illustrate the effectiveness of this method, a generalized nonlinear degradation model for the RUL prediction is built.

A. Feature Extraction

The degradation features that reveal the system's underlying health status can be extracted from the raw sensor signals. To obtain effective features, signal processing techniques, correlation analysis, or other statistical approaches are employed. Time-domain analysis, frequency-domain analysis, and time-frequency domain analysis are the common signal processing techniques for feature extraction [34]. Although the time-frequency domain has the best capacity for processing the nonstationary signals with noise, it requires high computational time and is not suitable for online prognostics. Moreover, the frequency domain analysis is mainly effective for anomaly detection and not suitable for RUL prediction. For the time domain analysis, it can be used for various systems because of

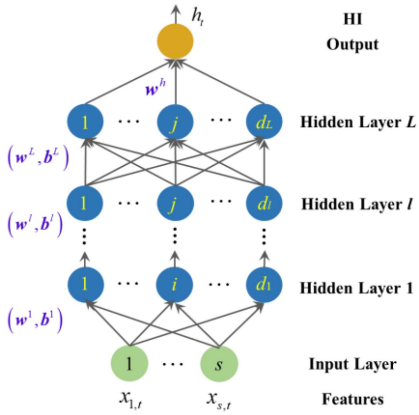

Fig. 1. Multilayer DNN architecture for feature fusion.

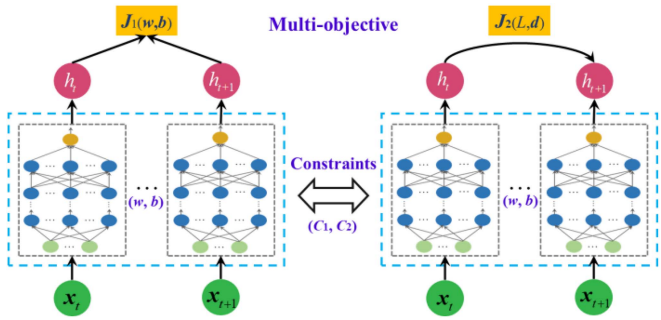

Fig. 2. Architecture of the DNN-based feature fusion model for HI construction.

the simplicity and fast computation time. Considering that this study aims to propose a generalized HI construction method, only the time domain analysis is investigated.

Since the raw monitoring signals are usually noisy due to the external interference, some pretreatment techniques, such as smoothing methods, filter algorithms, and singular spectrum analysis, are used for noise-reduction. Assume that s features related to the degradation process can be extracted from the noise-free signals. Let $x_{q,j}$ be the qth feature at t_j, $q = 1, \ldots, s$. As the features from different sources can have different scales, all features need to be standardized before the feature fusion.

B. DNN-Based Feature Fusion Architecture

In this section, we propose a DNN-based feature fusion model to construct a 1-D HI based on the extracted degradation features. A nonlinear mapping $\mathrm{f}: \boldsymbol{x} \in \mathbb{R}^s \rightarrow h \in \mathbb{R}$ from the feature space to the HI space is developed. Without loss of generality, Fig. 1 shows a multilayer DNN architecture with fully connected hidden layers. The input of DNN at time t_j is the s degradation features, i.e., $\boldsymbol{x}_j = [x_{1,j}, x_{2,j}, \ldots, x_{s,j}]^{\boldsymbol{T}}$. The output of DNN is the constructed HI.

Let $\boldsymbol{w}_i^l \in \mathbb{R}^{d_{l-1}}$ and b_i^l denote the weight and the bias between layer $l-1$ and the node i at layer l, respectively. The output of neuron i at the first hidden layer is $\eta_i^1 = \delta(\boldsymbol{w}_i^1 \boldsymbol{x}_j + b_i^1)$, where δ is the sigmoid activation function for nonlinear transformation. Furthermore, the output of the node i at layer l is denoted as $\eta_i^l = \delta(\boldsymbol{w}_i^l \boldsymbol{\eta}^{l-1} + b_i^l)$, where $\boldsymbol{\eta}^{l-1} \in \mathbb{R}^{d_{l-1}}$ is the input vector of layer l, $l = 2, \ldots, L$. Let $h_j = \delta(\boldsymbol{w}^h \boldsymbol{\eta}^L + b^h)$ be the output of DNN, where $\boldsymbol{w}^h \in \mathbb{R}^{d_L}$ denotes the weight vector between layer L and the output layer, $\boldsymbol{\eta}^L \in \mathbb{R}^{d_L}$ is the output of layer L and b^h is the bias of the output layer.

The next task is to determine the topological parameters and unknown coefficients of DNN. Different from many previous deep learning-based methods that took the HI construction as supervised regression problems, this work converts it to an indirect supervised optimization task with unlabeled features by plugging the desired properties into the DNN-based feature fusion model. In this way, some desirable properties acted as the domain knowledge can be involved in the HI construction step to enhance the interpretability. A possible formulation of this optimization problem is to define one objective function by constructing a representation performance indicator of HI based on the weighted combination of those properties [21]. Then, the parameters can be estimated by single-objective optimization.

As DNNs with longer depths and widths may have better nonlinear fitting capability, the spatial complexity of the fusion model would increase rapidly when the representation performance achieves the optimum. However, an overly complex model could lead to overfitting and time-consuming which might be not suitable for practical applications. In addition, as we can see in [8], [17], and [32], the structures of the DNN-based fusion models are selected artificially or determined by cross validation. The former may result in a nonoptimal fusion model, while the latter would increase the computational burden. To address this issue, the spatial complexity can be set as another objective function for the optimization task besides the representation performance. Then, the model structure can be determined through multiobjective optimization.

Note that, the two objectives based on representation performance and spatial complexity conflict with each other. To make tradeoff between two or more conflicting objectives, a multiobjective problem is formulated. Then, the network coefficients (weights and bias) and the topological parameters (number of layers and number of neurons for each layer) can be estimated by the multiobjective optimization simultaneously. The obtained DNN-based feature fusion model is expected to have simpler structure and generate acceptable HIs.

The architecture of the proposed DNN-based fusion model is displayed in Fig. 2. The network coefficients and topological parameters are the same throughout all the sub-DNNs.

C. Multiobjective Optimization for Model Training

In this section, a multiobjective optimization problem for training the DNN-based fusion model is formulated. The representation performance of HI is evaluated by a composite indicator based on the weighted combination of the desired properties. We consider monotonicity, trendability, robustness, and scale similarity in the formulation of objective functions for the MOM. The last property, i.e., consistency, will be realized by data truncation and normalization as stated in Section II-A. The MOM is widely used in the HI construction [35] and in

the model training for deep learning-based prognostics methods [36], [37], [38].

The first objective function of MOM for the representation performance is formulated as follows:

$$J_1(\boldsymbol{w},\boldsymbol{b}) = \lambda_1 \mathrm{Mon}(\boldsymbol{h}) + \lambda_2 \mathrm{Tre}(\boldsymbol{h}) + \lambda_3 \mathrm{Rob}(\boldsymbol{h}) + \lambda_4 \mathrm{Sca}(\boldsymbol{h}) \tag{7}$$

where $\boldsymbol{w} \in \mathbb{R}^{d_L+\sum_{l=1}^{L} d_l d_{l-1}}$ and $\boldsymbol{b} \in \mathbb{R}^{1+\sum_{l=1}^{L} d_l}$ are the vectors of weights and bias of the DNN-based fusion model. $\lambda_k > 0, k = 1, 2, 3, 4$ are the hyperparameters for balancing the importance of the four properties and $\sum_{k=1}^{4} \lambda_k = 1$. When there is no specific requirement, we can assign the same value to λ_k or determine λ_k by cross validation. If one specific property is preferred according to the engineering requirements, its corresponding λ_k is set to a larger value. For example, if the monotonicity is preferred for a more significant trend of HIs, λ_1 can be set to 0.5, which is larger than other $\lambda_{k\neq 1}$. Therefore, by tuning the hyperparameters, the representation performance of the DNN-base model can adapt to various scenarios.

If both the monotonicity and consistency approach to their maximums, the scale similarity based on the maximal and minimal values of all HIs would also tend to its maximum. The value range of J_1 is confined in [0, 1]. $J_1 = 1$ suggests that the constructed HIs achieve the best desired properties, i.e., the best representation performance. Since regularization can reduce the variance of solutions and avoid overfitting, L_2-norm regularization is employed here. Then, (7) becomes

$$J_1'(\boldsymbol{w},\boldsymbol{b}) = \lambda_1 \mathrm{Mon}(\boldsymbol{h}) + \lambda_2 \mathrm{Tre}(\boldsymbol{h}) + \lambda_3 \mathrm{Rob}(\boldsymbol{h}) + \lambda_4 \mathrm{Sca}(\boldsymbol{h}) - \lambda \|\boldsymbol{w}\|_2^2 \tag{8}$$

where λ is the tuning factor and usually set to be higher enough.

The spatial complexity of DNN can be measured by the topological parameters. Thus, the second objective function of the MOM for the spatial complexity is formulated as follows:

$$J_2(L,\boldsymbol{d}) = \sum_{l=1}^{L} d_l (d_{l-1}+1) + d_L + 1 \tag{9}$$

where $\boldsymbol{d} = [d_1, \ldots, d_L] \in \mathbb{R}^L$.

By combining (8) and (9), the formulation of MOM can be expressed as

$$\begin{aligned} &\max J_1'(\boldsymbol{w},\boldsymbol{b}), \min J_2(L,\boldsymbol{d}) \\ &\text{s.t. } L \in \mathbb{N} \geq 2, \boldsymbol{d} \in \mathbb{N}^L \\ &\boldsymbol{w} \in \mathbb{R}^{d_L+\sum_{l=1}^{L} d_l d_{l-1}}, \boldsymbol{b} \in \mathbb{R}^{1+\sum_{l=1}^{L} d_l}. \end{aligned} \tag{10}$$

The stochastic optimization algorithms are usually used to trained DNNs. However, these algorithms which minimize the expected loss function are mostly for single-objective optimization. Furthermore, since the values of L and $\boldsymbol{d}$ determine the dimensions of $\boldsymbol{w}$ and $\boldsymbol{b}$, the MOM in (10) is an uncertainty optimization problem. Hence, most common multiobjective optimization algorithms, such as strength pareto evolutionary algorithm, nondominated sorting genetic algorithm-II, multiobjective evolutionary algorithm, and multiple objective particle swarm optimization [39], [40], [41], are also not suitable for solving (10). To search the optimal solutions and train the DNN-based model more effectively and rapidly, we first define a cost-effectiveness ratio function as follows:

$$\max A(\boldsymbol{w},\boldsymbol{b},L,\boldsymbol{d}) = \mathbf{max}\frac{J_1'(\boldsymbol{w},\boldsymbol{b})}{J_2(L,\boldsymbol{d})} \tag{11}$$

where $A(\cdot)$ can be interpreted as the benefit of representation performance per unit spatial complexity. Thereupon, the MOM is reformulated to a single-objective optimization problem, which integrates two considerations. An HSADE algorithm is proposed for solving (11). Compared with the basic differential evolution algorithms, the HSADE has a less computational cost and a faster convergence property. The detailed procedure of HSADE is presented in Algorithm 1.

Algorithm 1: Framework of HSADE.

Input: The ranges of L and d_l, i.e., $[L^0, L^1]$ and $[d^0, d^1]$; N_P (population size).
Output: The optimal solutions $\boldsymbol{w}^*$, $\boldsymbol{b}^*$, L^* and $\boldsymbol{d}^*$.

1: **For** $L \in [L^0, L^1]$, $d_l \in [d^0, d^1]$**do**:
2: Set $g = 0$ and initialize a population $\boldsymbol{P}^g = \{\boldsymbol{\xi}_1^g, \cdots, \boldsymbol{\xi}_{NP}^g\}$ via uniform sampling, where $\boldsymbol{\xi}_i^g = [\boldsymbol{w}_i^g, \boldsymbol{b}_i^g]$.
3: **while** the termination rule is not satisfied **do**:
4: **for** $i = 1{:}\, N_P$
5: **Mutation**: Randomly pick three individuals $\boldsymbol{\xi}_{i,1}^g$, $\boldsymbol{\xi}_{i,2}^g$ and $\boldsymbol{\xi}_{i,3}^g$ from $\boldsymbol{P}^g$, and construct a new individual $\boldsymbol{\rho}_i^g$ for $\boldsymbol{\xi}_i^g$ $\boldsymbol{\rho}_i^g = \boldsymbol{\xi}_{i,1}^g + Sf_i^g \times (\boldsymbol{\xi}_{i,2}^g - \boldsymbol{\xi}_{i,3}^g)$, where scaling factor Sf_i^g is updated by Algorithm 2. $\boldsymbol{\xi}_{i,1}$ and $\boldsymbol{\xi}_{i,3}$ are the best and the worst individual in the three samples, respectively.
6: **Crossover**: Conduct binomial crossover between $\boldsymbol{\rho}_i^g$ and $\boldsymbol{\xi}_i^g$ to yeild a trial vector $\boldsymbol{U}_i^g = [u_{i,1}^g, u_{i,2}^g, \ldots, u_{i,D_L}^g]$ with $u_{i,j}^g = \begin{cases} \rho_{i,j}^g, \mathrm{rand}_j(0,1) \leq Cr_i^g or \mathrm{j} = j_{\mathrm{rand}} \\ \xi_{i,j}^g, \mathrm{otherwise} \end{cases}$ where crossover rate Cr_i^g is updated by Algorithm 2, and j_{rand} is an integer and randomly sampled from $\{1, 2, \cdots, D_L\}$. The dimension is $D_L = d_L + \sum_{l=1}^{L} d_l d_{l-1} + \sum_{l=1}^{L} d_l + 2 + L$.
7: **end for**
8: **Selection**: Sample two individuals from $\{\boldsymbol{\xi}_1^g, \cdots, \boldsymbol{\xi}_{NP}^g, \boldsymbol{U}_1^g, \cdots, \boldsymbol{U}_{NP}^g\}$ randomly, i.e., $\boldsymbol{\xi}_1'$ and $\boldsymbol{\xi}_2'$; the new individual is selected as $\boldsymbol{\xi}^{new} = \begin{cases} \xi_1', if \mathrm{A}(\xi_1') > \mathrm{A}(\xi_2') \\ \xi_2', \mathrm{otherwise} \end{cases}$
9: Repeat step 9 to obtain a new population $\boldsymbol{P}^{new}$. Set $g = g + 1$.
10: **end while**.
11: **end for**
12: $(\boldsymbol{w}^*, \boldsymbol{b}^*, L^*, \boldsymbol{d}^*) \leftarrow$ Return the optimal $(\boldsymbol{\xi}, L, \boldsymbol{d})^*$ with the best $A(\cdot)$

Algorithm 2: Self-adaptive Strategies for the Parameters in HSADE.

Input: Individual $\boldsymbol{\xi}_i^g$, $\boldsymbol{\xi}_{i,1}^g$, $\boldsymbol{\xi}_{i,2}^g$ and $\boldsymbol{\xi}_{i,3}^g$, Population $\boldsymbol{P}^g$.
Output: Sfig+1 and Crig+1
Self-adaptive updating of Sf
1: Calculate $A_{i,1} = A(\boldsymbol{\xi}_{i,1}^g)$, $A_{i,2} = A(\boldsymbol{\xi}_{i,2}^g)$ and $A_{i,3} = A(\boldsymbol{\xi}_{i,3}^g)$.
2: Calculate the multiplier $Ras = (A_{i,2} - A_{i,1})/(A_{i,3} - A_{i,1})$.
3: Update Sfig+1 by $Sf_i^{g+1} = Sf_l + (Sf_u - Sf_l) \times Ras$, where Sfu and Sfl denote the upper and the lower limits of the scaling factor, respectively.
Self-adaptive updating of Cr
5: Calculate $A_i = A(\boldsymbol{\xi}_i^g)$, $A_{\max} = \max_i A(\boldsymbol{\xi}_i^g)$, $A_{\min} = \min_i A(\boldsymbol{\xi}_i^g)$, and the average objective function $\bar{A} = \sum A(\boldsymbol{\xi}_i^g)/NP$.
6: Calculate the multiplier $Rac = (A_i - A_{\max})/(A_{\min} - A_{\max})$.
7: Update Crig+1 as follow
$$Cr_i^{g+1} = \begin{cases} Cr_l + (\mathrm{Cr}_u - Cr_l) \times Rac, A_i \le \bar{A} \\ Cr_l, A_i > \bar{A} \end{cases},$$
where $\mathrm{Cr_u}$ and $\mathrm{Cr_l}$ are respective the upper limit and lower limit of the crossover rate

In order to maintain a better balance between the proximity and diversity of the individuals, the self-adaptive strategies of scaling factor and crossover rate in HSADE are developed by simultaneously utilizing prior and posteriori knowledges in the evolutionary process, which are presented in Algorithm 2.

The training data from historical units should be run-to-failure data, so that the scale similarity and consistency properties can be formulated. By substituting the extracted feature data into the DNN-based fusion model, the optimal HI can be obtained through the MOM and HSADE. Note that, given specific values of L and $\boldsymbol{d}$, the dimensions of $\boldsymbol{w}$ and $\boldsymbol{b}$ are determined and their estimates can be searched by the Steps 2–12 of HSADE. When traversing L and d_l in their value ranges, we can obtain a set of candidate solutions for $\boldsymbol{w}$ and $\boldsymbol{b}$. Then, the final optimal network coefficients and topological parameters with the best cost-effectiveness ratio are selected from the candidate solution sets. This solving process is obviously a multihierarchical strategy, which is the reason for naming the proposed algorithm as hierarchical SADE. After the training process, uncomplete degradation data of the testing units is substituted into the trained DNN-based fusion model, the HI sequences of the testing units can be achieved.

D. RUL Prediction

The constructed HIs are substituted into a degradation model for prognostics. WP [42], [43] is commonly used as a degradation model to describe the failure propagation, i.e.,

$$h(t) = h_{\mathbf{0}} + \mu(t;\boldsymbol{\theta}) + \sigma_B B(t) \tag{12}$$

where $\mu(t;\boldsymbol{\theta})$ is the drift function, $\boldsymbol{\theta}$ is the parameter vector, σ_B is the diffusion parameter, and $B(t)$ is the standard Brownian motion. For convenience, the initial HI h_0 is set to 0, i.e., $h_0 = 0$. If the initial HI is not equal to 0, h_0 can be subtracted from $h(t)$ at the outset. To describe the nonlinearity of the degradation processes, the power-law form and the exponential law form are commonly used to formulate $\mu(t;\theta)$, i.e.,

$$\mu(t;\boldsymbol{\theta}) = \begin{cases} \alpha_1 t^{\alpha_2} \\ \alpha_1 \exp(\alpha_2 t) \end{cases}. \tag{13}$$

Although (13) can model an enormous number of common nonlinear degradation trajectories, they might struggle to describe some complex degradation processes which are both nonconvex and nonconcave. To improve the generality and flexibility of degradation modeling without prior information, a novel method is developed. The basic idea is to introduce the kernel tricks into the drift function as follows:

$$\mu(t;\boldsymbol{\theta}) = \sum_{k=1}^{m} \theta_k \kappa(t, t_k) + \theta_0 \tag{14}$$

where $\kappa(t, t_k)$ is the kernel function, θ_k is the weight, θ_0 is the bias, m is the sample size, and $\boldsymbol{\theta} = [\theta_0, \theta_1, \ldots, \theta_m]^{\boldsymbol{T}}$. The Gaussian kernel function is usually selected in nonlinear model prediction [44]. Then, a generalized nonlinear Wiener process degradation (GNWP) model is formulated by combining (12) and (14) as follows

$$h(t) = h_{\mathbf{0}} + \mu(t;\boldsymbol{\theta}) + \sigma_B B(t) \cdot \mu(t;\boldsymbol{\theta}) = \sum_{k=1}^{m} \theta_k \kappa(t, t_k) + \theta_0. \tag{15}$$

The degradation increments of His at t_j are denoted as

$$\Delta h_j = h(t_j) - h(t_{\mathrm{j-1}}) = \sum_{k=1}^{m} \theta_k \boldsymbol{\Delta}\kappa_{j,k} + \sigma_B \Delta B_j \tag{16}$$

where $\boldsymbol{\Delta}\kappa_{j,k} = \kappa(t_j, t_k) - \kappa(t_{j-1}, t_k)$ and $\boldsymbol{\Delta}B_j = B(t_j) - B(t_{j-1})$. Based on the properties of WP, the increment $\boldsymbol{\Delta}\mathrm{h}_j$ follows a Gaussian distribution. Assume the sensor signals are observed with equal sampling interval, i.e., $\tau = t_j - t_{j-1}$, $j = 1, 2, \ldots, m$. Then, the likelihood function of the complete data $\boldsymbol{\Delta}\mathbf{h}$ can be given by

$$p(\boldsymbol{\Delta h}\,|\boldsymbol{\theta}, \sigma_B) = \left(2\pi\sigma_B^2\tau\right)^{-\frac{m}{2}} \exp\left(-\frac{\|\boldsymbol{\Delta h} - \boldsymbol{\Phi\theta}\|^2}{2\sigma_B^2\tau}\right) \tag{17}$$

where $\boldsymbol{\Delta h} = [\Delta h_1, \ldots, \Delta h_m]^{\boldsymbol{T}}$, $\boldsymbol{\Phi}_{m\times(m+1)} = [\phi(t_1), \ldots, \phi(t_m)]^{\boldsymbol{T}}$, and $\phi(t_j) = [1, \boldsymbol{\Delta}\kappa_{j,1}, \ldots, \boldsymbol{\Delta}\kappa_{j,m}]$.To implement the parameter estimation, a zero-mean Gaussian prior is defined over θ as

$$p(\boldsymbol{\theta}\,|\boldsymbol{\beta}) = \prod_{k=0}^{m} \sqrt{\frac{\beta_k}{2\pi}} \exp\left(-\frac{\theta_k^2\beta_k}{2}\right) \tag{18}$$

Where $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_m]^{\boldsymbol{T}}$ is the hyperparameter vector. Based on the conjugate distribution, the posterior distribution

Algorithm 3: RUL Prediction.

Input: History of the HI values up to time t_j, $\boldsymbol{h}_{1:j}=[h_1,\cdots,h_j]^{\mathrm{T}}$.
Output: The PDF of the RUL at t_j, $\mathrm{f}_{R_j}(r)$.
1: Select the kernel function in (14).
2: Calculate the posterior estimate of θ and the estimate of σ_B based on the HI values and the method presented in Appendix A.
3: Predict the degradation increment Δh_{j+1} at time t_{j+1} by (A4).
4: Calculate the PDF of the RUL at t_j, $\mathrm{f}_{R_j}(r|\boldsymbol{\theta},\sigma_B)$ by (22), which θ is replaced by its posterior mean, i.e., $\mathrm{f}_{R_j}(r)=\mathrm{f}_{R_j}(r|\boldsymbol{\theta},\sigma_B)|_{\boldsymbol{\theta}=\mathbf{v}}$.
5: Set $j=j+1$ and return to step 2.

of θ is derived as follows:

$$p\left(\boldsymbol{\theta}\left|\boldsymbol{\Delta h},\boldsymbol{\beta},\sigma_B\right.\right)=(2\pi)^{-\frac{m+1}{2}}\left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}}\times\exp\left[-\frac{(\boldsymbol{\theta}-\boldsymbol{v})^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\boldsymbol{v})}{2}\right] \tag{19}$$

where the posterior mean and covariance are, respectively

$$\boldsymbol{v}=\sigma_B^{-2}\tau^{-1}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{\boldsymbol{T}}\boldsymbol{\Delta h},\boldsymbol{\Sigma}=\left(\sigma_B^{-2}\tau^{-1}\boldsymbol{\Phi}^{\boldsymbol{T}}\boldsymbol{\Phi}+\boldsymbol{\Lambda}\right)^{-1}$$

where $\boldsymbol{\Lambda}$=$diag(\boldsymbol{\beta})$. The optimal estimates of θ and σ_B can be calculated by maximizing the marginal likelihood in (A1). The details for parameter estimation are given in Appendix A.

The lifetime T is generally defined as the FPT of $\mathrm{h}(t)$ exceeding a prefixed threshold D_f , i.e.,

$$T=\inf\left\{t:h\left(t\right)\geq D_f\left|h_0<D_f\right.\right\}. \tag{20}$$

According to (20), the RUL at t_j can be defined as the residual time for the HI value exceeding D_f from h_j, i.e.,

$$R_j=\inf\left\{\mathrm{r}:h\left(t_j+r\right)\geq D_f\left|h_j<D_f\right.\right\}. \tag{21}$$

Under the assumption in Section III-C, $D_f=1$.

From the results in [27], given the HI h_j, the probability density function (PDF) of R_j for GNWP is approximated with an explicit form as

$$f_{R_j}\left(r\left|\boldsymbol{\theta},\sigma_B\right.\right)\cong\frac{1}{\sqrt{2\pi r}}\exp\left\{-\frac{S_j^2\left(r\right)}{2r}\right\}\times\left(\frac{S_j\left(r\right)}{r}+\frac{1}{\sigma_B}\frac{d\mu\left(t_j+r;\boldsymbol{\theta}\right)}{dr}\right) \tag{22}$$

where $S_j(r)=(D_f-h_j-\mu(t_j+r;\boldsymbol{\theta})+\mu(t_j;\boldsymbol{\theta}))/\sigma_B$.

The procedure of the RUL prediction is summarized in Algorithm 3. Moreover, we elaborate the flowchart of implementing the HI construction and RUL prediction in Fig. 3.

IV. Case Studies

This section presents two case studies to illustrate the superiorities of the proposed framework.

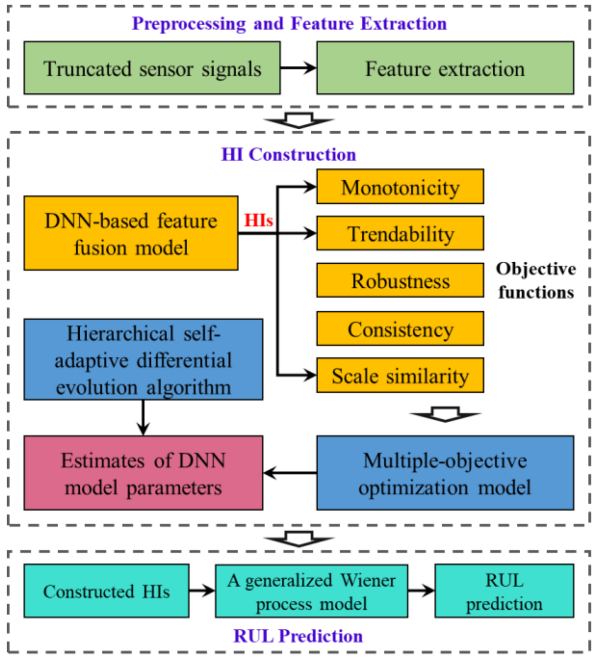

Fig. 3. Flowchart of implementing the HI construction and RUL prediction.

A. Lithium-Ion Battery Datasets

The lithium-ion battery datasets were from the NASA Ames Prognostics Center of Excellence [45]. The battery charge–discharge experiments were conducted to collect the health parameters of batteries. The signals of temperature, capacity, voltage, and current, etc., were measured with the running cycles. The failure of a battery is defined to occur once its capacity declines below 1.4 Ah. Four batteries #5, #6, #7, and #18 are selected in this case, where batteries #6 and #18 are used as training samples and the other two are used as testing samples.

Most existing studies focused on using the capacity as the HI for the battery RUL prediction [46]. However, considering that it could be difficult to measure the capacity in some online scenarios and directly monitoring some internal state information of batteries through regular sensors may be expensive, it is more practical to use the common monitoring data of batteries for extracting some effective features that can depict the battery performance. Hence, we extract four features from the raw sensor data. The definitions of these four features can be referred to [30]. The battery capacity is normalized with the initial value and the failure threshold, which is called as state-of-health (SoH). The SoH is usually taken as the actual HI of batteries. Then, we make comparisons between the SoHs and the constructed HIs to validate the proposed method.

After obtaining the degradation features, all these feature data are first truncated before the batteries approach the failure threshold, and second normalized to ensure the value ranges of all features fall within [0, 1], i.e.,

$$x'_{\mathrm{q},j}=\frac{x_{q,j}-x_{q,\min}}{x_{q,\max}-x_{q,\min}} \tag{23}$$

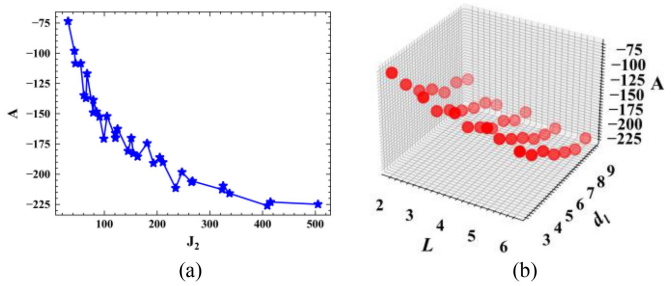

Fig. 4. Optimization results of the parameter estimation for the DNN-based feature fusion model. (a) Cost-effectiveness ratio versus the spatial complexity. (b) Representation performance versus the topological parameters.

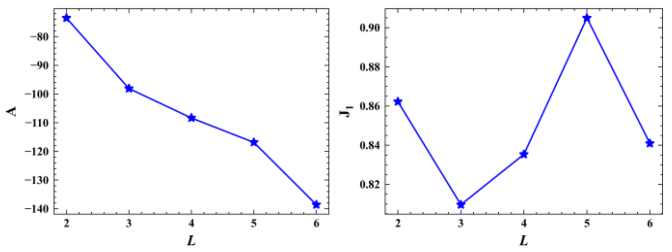

Fig. 5. Cost-effectiveness ratio A and representation performance J_1 under different values of L where $d_l = 3$.

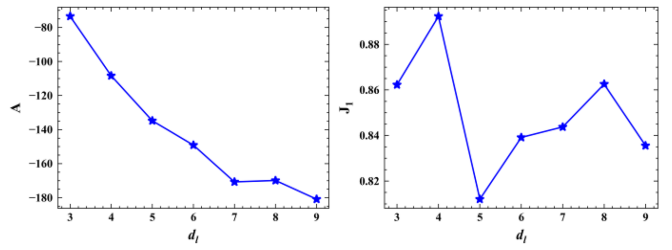

Fig. 6. Cost-effectiveness ratio A and representation performance J_1 under different values of d_l where $L = 2$.

where $x_{q,\max}$ and $x_{q,\min}$ are the highest and the lowest values of the qth feature, respectively. Then, the preprocessed feature data are substituted into the proposed method to generate composite HIs. The hyperparameters λ_1, λ_2, λ_3, and λ_4 in (7) are assumed to be 0.25, 0.25, 0.25, and 0.25 respectively. Based on the feature datasets of batteries #5 and #6, the optimization results of the parameter estimation for the DNN-based model by Algorithm 1 are shown in Fig. 4. The cost-effectiveness ratio function declines as the spatial complexity increases in Fig. 4(a). This suggests that a simpler DNN-based fusion model is preferred. This conclusion is also supported by Fig. 4(b). It can be found that when the topological parameters of the DNN-based feature fusion model are set as $L = 2$ and $d_l = 3$, the cost-effectiveness ratios achieve its maximum. Under this configuration, the resulted model can have a higher representation performance with a lower spatial complexity.

Furthermore, we explore the optimal results of the MOM under different topological parameter configurations. It can be found from Figs. 5 and 6 that the effectiveness ratios decline as L and d_l increase, which indicates the DNN-based fusion model with simpler structure is preferred. Both the representation performance indicators increase first and then decrease as L and d_l increase. That is, the optimal solution is a compromise between these two objective functions.

To illustrate the superiority of the proposed method, other two HI construction methods are used for comparison. One is adopting the Box–Cox transformation technique to extract the HIs (HI-BC) from a single feature (discharging voltage difference of equal time interval) of lithium-ion batteries [47]. The other one linearly fused the four features of batteries presented in Table I to obtain the composite HIs (HI-LN) by a single-objective optimization, whose objective function was formulated based on three desired properties [21].

Fig. 7 displays the training results of the DNN-based fusion model based on the datasets of batteries #6 and #18. HIC denotes the composite HIs constructed by the proposed method. The actual SoH is set as a benchmark. It can be seen from Fig. 7 that both the constructed composite HIs for two batteries perform as well as the actual SoHs, which verifies the proposed method. Then, the trained DNN model is applied to the testing datasets. Fig. 8 shows the testing results of the constructed HIs by different methods. Obviously, both the trends and values of HICs are closer to the SoHs than those of HI-BCs and HI-LNs. By comparing the HICs with HI-BCs, we can conclude that more degradation information is included in more features and thus the constructed HIs can reveal the underlying health status of batteries better. Through the comparison of HICs and HI-LNs, it can be ascertained that a more complex model with higher nonlinear fitting ability can capture both the trends and the fluctuations of heath states more precisely than the linear feature fusion model.

Additionally, to further compare those different HI construction methods, the absolute errors between the values of actual

TABLE I DEFINITION OF THE FEATURES FOR LITHIUM-ION BATTERY

Feature	Definition
CTV	Charging time interval of equal voltage difference from 3.8 V to 4.2 V
CTC	Charging time interval of equal current different from 1.48 A to 0.3 A
DTV	Discharging time interval of equal voltage difference from 3.7 V to 2.7 V
DVT	Discharging voltage difference of equal time interval

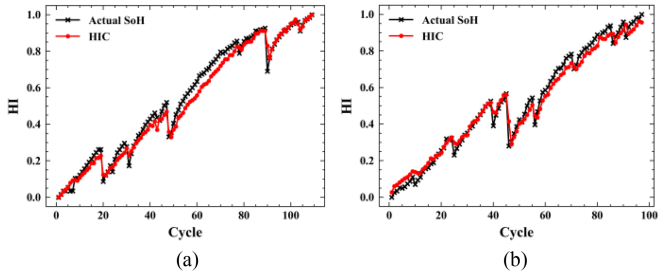

Fig. 7. Training results of the DNN-based feature fusion model based on the training datasets. (a) #6. (b) #18.

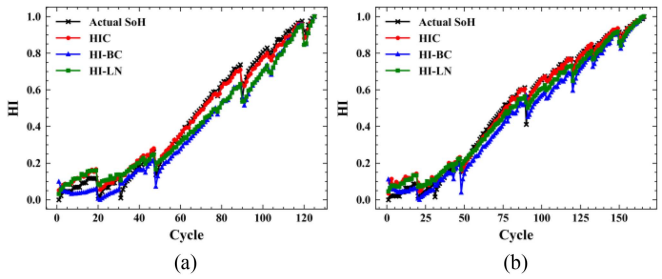

Fig. 8. Comparisons of the HIs constructed by different methods for the testing battery datasets. (a) #5. (b) #7.

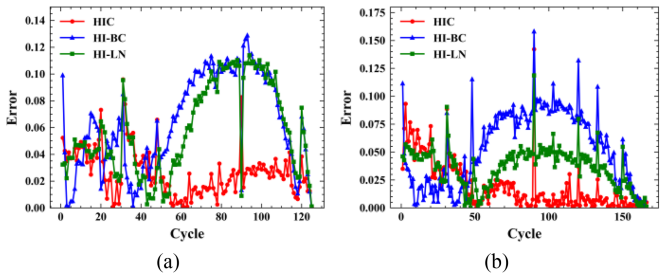

Fig. 9. Comparisons of the errors between the actual SoHs and the HIs constructed by three different methods for the testing battery datasets. (a) #5. (b) #7.

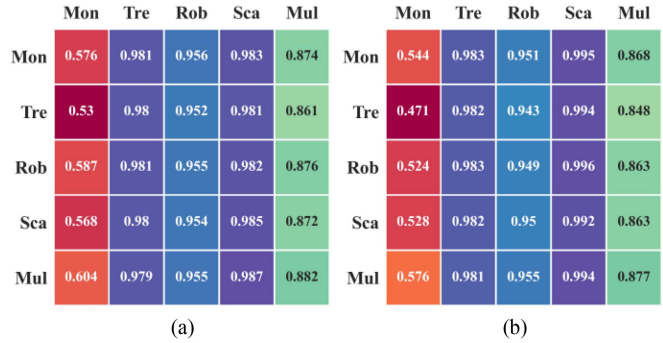

Fig. 10. Representation Performance of the constructed HIs under different combinations of the desired properties. (a) Training dataset. (b) Testing dataset.

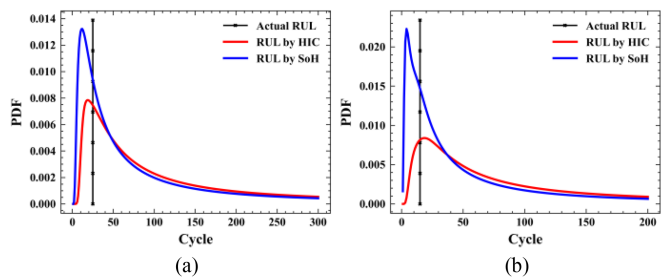

Fig. 11. Comparisons of the RUL predictions for battery #5 by the constructed HIs and the actual SoHs. (a) at cycles 100. (b) at cycles 110.

SoHs and constructed HIs for batteries #5 and #7 are calculated. Fig. 9 presents the corresponding results. At the early cycles, the errors of the three methods are close. While at the later cycles, the errors of HICs are significantly lower than those of HI-BCs and HI-LNs. Therefore, as more and more monitoring data are obtained, the proposed method outperform other methods more significantly. Furthermore, Table I presents the RMSEs between the actual SoHs and the values of constructed HIs by different methods for batteries #5 and #7. The comparative results indicate the proposed HI construction method achieves the smallest RMSE values for both batteries than other methods.

Besides validation of the proposed method, this section also analyzes whether the multiple desired properties are necessary to construct the HI or one simple property is enough. To generally investigate the effects of multiple desired properties, we first use only one property to formulate the objective function of MOM for the HI construction. Then, the representation performance of constructed HIs based on this property is evaluated through the other properties and also by the weighted sum of all properties (Mul) as shown in (7). Fig. 10 shows the HI performance based on the property given in the corresponding row, which is evaluated through individual property or multiple properties presented in columns. For example, in the first row of Fig. 10(a), the HIs based on the Mon property have the highest score of trendability on both two datasets. From Fig. 10(a) and (b), the HIs based on only one property, such as monotonicity or robustness, can also have high scores of the other properties. Besides, when using the combination of all the properties, the constructed HIs have the highest scores of the Mon property and the Mul, and achieve excellent performance in Tre, Rob, and Sca properties for both two datasets. Therefore, it is worth considering multiple desired properties in the HI construction, and thereby more monotonous and accurate HIs can be generated for prognostics.

Fig. 11 further validates the performance of the constructed HIs for the RUL prediction. The predicted PDFs of the RUL at different starting timepoints by the constructed HIs and by the SoHs for battery #5 are shown. According to the comparisons, both the predicted distributions cover the actual RULs well. Although the PDF peaks based on the SoHs are higher than those by the constructed HIs, the mean values of the former are far away from the actual RULs more than those of the latter. This indicates that the proposed HI construction method can generate HIs which represent the underlying health status effectively, when the internal health status is unavailable.

B. Rolling Bearing Datasets

The rolling bearing datasets are collected from PRONOSTIA platforms [48]. The datasets of Bearings 1-1 and 1-2 (run-to-failure data) are selected for model training, and the datasets of Bearings 1-3 and 1-4 (truncated data) are used for model testing. There are 12 time domain features extracted from the raw horizontal vibration signals, whose calculational formulas are listed in Table II. The last feature SF is used to extract the signal trend and remove signal noises with a polynomial fitting function. Since the RUL prediction is a long-term task, the time window is set to 10 s, which means all the features are extracted every 10 s.

TABLE II
RMSES BETWEEN THE VALUES OF CONSTRUCTED HIS AND ACTUAL SOHS FOR TWO BATTERIES

Methods	HIC	HI-BC	HI-LN
#5	0.0318	0.0729	0.0673
#7	0.0297	0.0638	0.0405

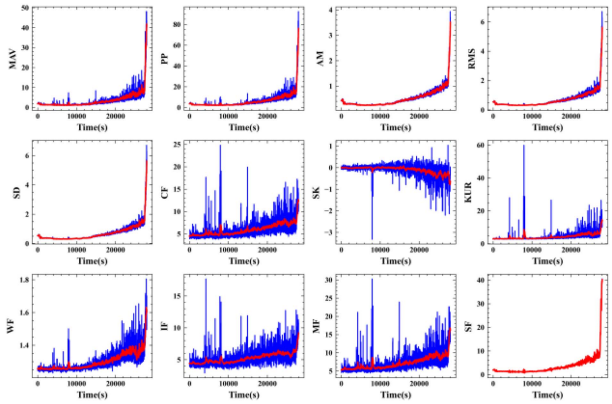

Fig. 12. Illustration of extracted degradation features and their trends of Bearing 1-1.

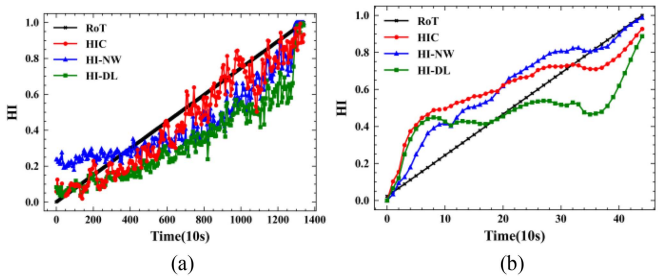

Fig. 13. Comparative training results of the HIs constructed by different methods. (a) Bearing 1-1. (b) Bearing 1-2.

Exponential smooth methods are used to remove the noise in all twelve extracted features. Fig. 12 shows the extracted degradation features for Bearing 1-1. The trends of most features are conspicuous and approximately monotonic, and present two distinct stages (normal stage and degradation stage). To improve the efficiency and accuracy of the HI construction and RUL prediction, only the datapoints of features at the degradation stage are selected. According to [7], the IDPs of Bearings 1-1and 1-2 are 14620 s and 8260 s, respectively. To eliminate the influence of the dimensions of different features, all the truncated feature data are Z-score standardized. The hyperparameters λ_1, λ_2, λ_3, and λ_4 in (7) are also assumed to be 0.25, 0.25, 0.25, and 0.25 respectively.

The preprocessed training datasets are substituted into the DNN-based fusion model and the unknown parameters are estimated by the MOM and Algorithm 1. Then, the topological parameters are determined as $L = 2$ and $d_l = 6$, respectively. Fig. 13 shows the HIs (HIC) of Bearings 1-1 and 1-2 constructed by the proposed method. Time 0 denotes the IDP. To

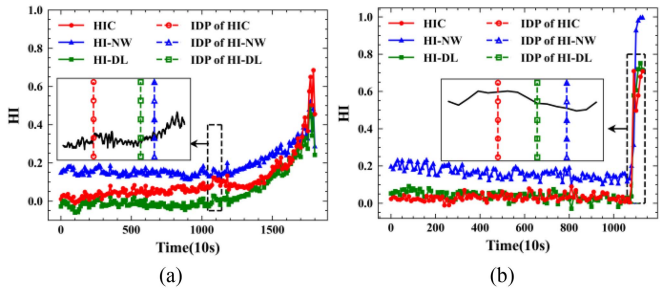

Fig. 14. Comparative results of the constructed HI and IDP detection by different methods for (a) Bearing 1-3. (b) Bearing 1-4.

demonstrate the superiority of the proposed method, we conduct a comprehensive comparison between different methods. Chen et al. [30] considered the properties of monotonicity, trendability, and robustness for HI construction (HI-NW) with a nonlinear weighted sum of multiple features. A deep learning-based data fusion approach was developed in [27] for HI construction (HI-DL) with only two desired properties (monotonicity and range information), whose topologies were determined through cross-validation and parameters were estimated by single-objective optimization. Note that the actual health states or HIs of bearings are always unknown. Some researchers used the reciprocal of failure time (RoT) as the underlying HI. However, the RoT only considers the monotonicity property and ignores the nonlinearity and stochasticity of degradation. The comparative HI results by different methods are presented in Fig. 13. All these methods adopt the same feature sequences.

From Fig. 13, we can observe that the HICs overall perform best on the training datasets when compared with the HI-NWs and HI-DLs. The HICs have more remarkable monotonic trends and can covers the RoTs better than HI-NWs or HI-DLs. This is possibly because the HI-NW just uses a nonlinear weighted sum of multiple features to construct the HI, which cannot capture complicated relationships among features. Since only two desired properties are considered and some vital information is ignored, the HI-DL does not perform as well as the HIC. Moreover, the nonlinear characteristics of degradation, such as stochasticity and volatility, are also captured by HICs. This indicates the proposed method accounting for multiple desired properties can be efficient in the prognostic tasks.

The datasets of Bearings 1-3 and 1-4 are then used to test the trained DNN-based fusion model. Fig. 14 shows the HIs constructed by different methods for these two bearings. Note that, all these HI trajectories present a normal stage and a degradation stage. The proposed method cannot only distinguish the two different heath stages accurately, but also generate effective HIs for prognostics. To compare the separability of different HI trajectories, all their IDPs are calculated by using the 3σ-criterion [7]. Since the IDPs of Bearings 1-1 and 1-2 in the training datasets are determined according to [7], the IDPs of Bearings 1-3 and 1-4, which were obtained based on RMS in [7], are taken as benchmarks for model testing. Table III shows the comparative results. It can be found that both the IDPs of HIC for Bearings 1-3 and 1-4 are slightly earliest than those of

TABLE III
TIME DOMAIN FEATURES OF VIBRATION SIGNALS

Feature	Formulation
Maximum absolute value (MAV)	$x_{MAV}=\max(\lvert z_1\rvert,\lvert z_2\rvert,\cdots,\lvert z_L\rvert)$
Peak to Peak value (PP)	$x_{PP}=\lvert\max(z_l)-\min(z_l)\rvert$
Root mean square (RMS)	$x_{RMS}=\sqrt{\frac{1}{L}\sum_l z_l^2}$
Absolute mean (AM)	$x_{AM}=\frac{1}{L}\sum_l\lvert z_l\rvert$
Standard deviation (SD)	$x_{SD}=\sqrt{\frac{1}{L-1}\sum_l(z_l-\bar{z})^2}$
Wave factor (WF)	$x_{WF}=x_{RMS}/x_{AM}$
Crest factor (CF)	$x_{CF}=x_{MAV}/x_{RMS}$
Impulse factor (IF)	$x_{IF}=x_{MAV}/x_{AM}$
Margin factor (MF)	$x_{MF}=x_{MAV}/x_{SMR}$
Skewness (SK)	$x_{SK}=\frac{1}{(L-1)x_{SD}^3}\sum_l(z_l-\bar{z})^3$
Kurtosis (KUR)	$x_{KUR}=\frac{1}{(L-1)x_{SD}^4}\sum_l(z_l-\bar{z})^4$
Smooth function (SF)	$x_{SF}=polyfit\,[(z_l)]$

Remark: $\bar{z}$ is the average value of the raw signal $z_1, z_2, \cdots, z_L$.

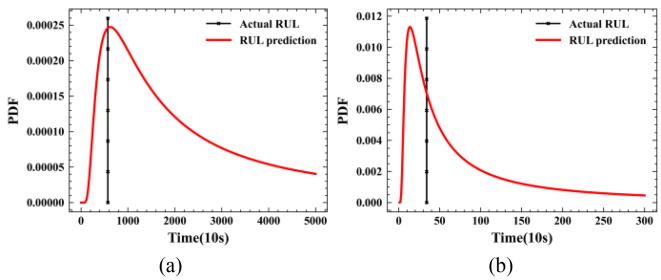

Fig. 15. Predicted RUL distriutions for (a) Bearing 1-3, (b) Bearing 1-4.

HI-NW and HI-DL. The IDPs of HIC and RMS for Bearing 1-4 are close, while the IDP of HIC for Bearing 1-3 is significantly earlier than that of RMS. It suggests that the HIs by the proposed method have the highest sensitivity and generality and the most notable separability. Hence, we can conclude that the occurrence of degradation of different bearings can be identified timely by the proposed method even if they are in the incipient failure stages.

Next, the HI sequences after IDPs are input into GNWP for the RUL prediction. The start timepoints of bearings 1-3 and 1-4 for prognostics are 18010 s and 11380 s, respectively. When the value of HI crosses 1, the bearing fails. The RUL predictions are displayed in Fig. 15, which verify the proposed GNWP model based on HICs. The actual RULs close to the peaks of the PDFs can be well covered by the predicted distributions. It indicates that the proposed method can be not only used to characterize the degradation process effectively, but also provide accurate uncertain quantification for the RULs.

Furthermore, some existing methods are selected for comparisons with the evaluation metric summarized in Section II-B. A WP degradation model is used to predict the RUL distributions

TABLE IV
DETECTION OF THE IDPS OF THE CONSTRUCTED HIS BY DIFFERENT METHODS FOR TESTING BEARING DATASETS

Methods	HIC	HI-NW	HI-DL	RMS
Bearing 1-3	10750 s	14250 s	13460 s	13650 s
Bearing 1-4	10920 s	10990 s	10960 s	10840 s

TABLE V
COMPARATIONS OF THE RUL PREDICTIONS BY DIFFERENT METHODS FOR THE TESTING BEARING DATASETS

Metrics/Methods	HIC+GNWP	HI-NW+WP	HI-DL+ Poly
MARE	0.3378	0.3383	0.500
RMSE	38.079	84.202	191.343
MAD	49.498	72.125	178.190
PS	0.3770	0.4436	1.476
Metrics/Methods	MCNN	RNN	GRU
MARE	0.2049	0.5546	0.0745
RMSE	70.937	176.115	13.038
MAD	65.054	159.100	9.899
PS	0.3448	1.2887	0.0572

based on the HI-NWs in [30], and a polynomial (Poly) degradation model is adopted to fit the HI-DLs for RUL prediction in [27]. As many prognostics methods with end-to-end manner were studied in the recent years, some recent deep learning-based methods are also selected for comparison, such as multiscale convolutional neural network (MCNN) [50], recurrent neural network (RNN), and gated recurrent unit (GRU) [10] are used here to predict the RULs, where multiple degradation features and the RUL percentages are used as the input and output, respectively. The performance evaluation metrics including MARE, RMSE, MAD, and PS of all these prognostic methods for the testing datasets are calculated and the results are presented in Table IV. As some predictions are in the form of distributions, the maximum probable values are adopted as the predicted RUL values.

We can intuitively find from Table V that the RMSEs and MADs of the "HI-DL+Poly" method are significantly higher than those of other two HI-based methods. It verifies that, due to considering more desired properties, the proposed method and the "HI-NW+WP" method perform better in prognostics. Another reason may be that the polynomial model without the consideration of the degradation stochasticity has no capacity for describing complicated degradation processes and yielding higher-accuracy predictions. Since the absolute RUL values are larger, the differences among the MAREs of different methods accounting for the relative errors between the predicted and actual values are smaller. Moreover, the PS values of the "HIC+GNWP" method and the "HI-NW+WP" method are close, both of which are far less than that of the "HI-DL+Poly" method. This indicates the predicted RULs by the former two methods are relatively ahead of those by the last method. Note that we generally prefer proper earlier RUL predictions and making maintenance actions in advance to avoid unnecessary failure loss. Moreover,

by comparing the proposed method with the end-to-end models, we find that the proposed "HIC+GNWP" method performs better than RNN, compare beauty with MCNN, and slightly inferior to GRU. However, since the end-to-end prognostics methods always only provide the point estimates of RUL and do not consider domain knowledge, the prediction task act as a black-box, which cannot reflect the evolution of failure propagation and cannot give a straightforward visualization of the nonlinear degradation process. Moreover, the uncertainty quantification that is an important issue in practice is also ignored in those end-to-end methods. Therefore, the proposed method which not only depicts the underlying degradation status clearly, but also predicts the RUL distributions with analytic forms, can have better generalization ability and efficiency to model the underlying degradation mechanisms of various systems under different scenarios. Then, prescriptive decisions can be made more flexibly according to the engineering requirements.

V. Conclusion

This article has developed a feature fusion-based HI construction method with deep learning and multiobjective optimization. The extracted features were used to construct the composite HI by a DNN-based feature fusion model. Multiple desired properties for the practical requirements of prognostics were considered as the basis of the objective function in the DNN training. To make a compromise between the representation performance and the spatial complexity of DNN, a MOM is developed. Then, the resulted HIs were input into a generalized nonlinear WP model with kernel functions for the RUL prediction. The comparisons with other methods in the two illustrative examples demonstrate the effectiveness and applicability of the proposed method.

Although the proposed method performs well in prognostics, some future studies are worthy of investigating. The first one is how to deal with nonmonotonic degradation signals under dynamic operating environment. The other one is to extend the HI construction to early fault detection.

Appendix A

The marginal log-likelihood of θ is calculated by

$$\begin{aligned}\ln p\left(\Delta \boldsymbol{h}\,|\boldsymbol{\beta},\sigma_B\right) &= \ln \int p\left(\Delta \boldsymbol{h}\,|\boldsymbol{\theta},\sigma_B\right) p\left(\boldsymbol{\theta}\,|\boldsymbol{\beta}\right) d\boldsymbol{\theta} \\ &= -\frac{1}{2}\left\{m\ln\left(2\pi\right)+\ln|\boldsymbol{\Psi}| \right. \\ &\quad \left. +\,\Delta \boldsymbol{h}^{\boldsymbol{T}}\boldsymbol{\Psi}^{-1}\Delta \boldsymbol{h}\right\} \end{aligned} \tag{A1}$$

where $\boldsymbol{\Psi}=\sigma_B^2\tau\mathbf{I}+\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Phi}^{\boldsymbol{T}}$. Note that it is hard to calculate the estimates of β and σ_B by maximizing (A1) in the closed form. Hence, an iterative estimation method is proposed as follows.

Calculating the first partial derivatives of (A1) with respect to β and σ_B yields

$$\beta_k^{new}=\frac{1-\beta_k\boldsymbol{\Sigma}_{k,k}}{v_k^2} \tag{A2}$$

$$\left(\sigma_B^2\right)^{\text{new}}=\frac{\|\Delta \boldsymbol{h}-\boldsymbol{\Phi}\boldsymbol{\nu}\|^2}{(m+1)-\sum_{k=0}^{m}\left(1-\beta_k\boldsymbol{\Sigma}_{k,k}\right)} \tag{A3}$$

where v_k is the kth element of the posterior mean $\mathbf{v}$, and $\boldsymbol{\Sigma}_{k,k}$ is the kth diagonal element of the posterior covariance $\boldsymbol{\Sigma}$ in (19), which are calculated with the current β and σ_B.

The procedure of the iterative estimation are as follows.

Step 1: Initialize β and σ_B.
Step 2: Calculate the posterior distribution of θ by (19).
Step 3: Re-estimate β_k^{new} and $\left(\sigma_B^2\right)^{new}$ by (A2) and (A3).
Step 4: Repeat Steps 2–3 until convergence.

As a result, the optimal estimates of θ and σ_B can be obtained, which are denoted as $\boldsymbol{\theta}^*$ and σ_B^*. According to [49], the iterative estimation is equivalent to an expectation–maximization update and so is guaranteed to locally maximize the marginal log-likelihood in (A1). Utilizing (A2), the iterative process results in faster convergence and we would not encounter any optimization difficulties in practice [49].

When given a new timepoint t_{m+1}, the posterior distribution of $\boldsymbol{\Delta}h_m$ is predicted with the estimated results of $\boldsymbol{\theta}^*$ and σ_B^*. Note that the posterior distribution of $\boldsymbol{\Delta}h_m$ is also a Gaussian

$$\begin{aligned} p\left(\Delta h_{m+1}\,|\Delta \boldsymbol{h},\boldsymbol{\theta}^*,\sigma_B^*\right) &\sim \mathcal{N}\left(\boldsymbol{v}^{*\boldsymbol{T}}\phi\left(t_{m+1}\right),\sigma_B^{2*}\tau\right. \\ &\quad \left.+\,\phi(t_{m+1})^{\boldsymbol{T}}\boldsymbol{\Sigma}^*\phi\left(t_{m+1}\right)\right) \end{aligned} \tag{A4}$$

where $\boldsymbol{v}^*$ and $\boldsymbol{\Sigma}^*$ are the posterior mean and covariance of $\boldsymbol{\theta}^*$.

Acknowledgment

The authors would like to thank the anonymous referees for their remarkable comments.

References

[1] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, May 2018.
[2] T. Wang, Z. Liu, M. Liao, N. Mrad, and G. Lu, "Probabilistic analysis for remaining useful life prediction and reliability assessment," *IEEE Trans. Rel.*, vol. 71, no. 3, pp. 1207–1218, Sep. 2022, doi: 10.1109/TR.2020.3032157.
[3] N. Gebraeel, M. Lawley, R. Li, and J. Ryan, "Residual-life distributions from component degradation signals: A Bayesian approach," *IIE Trans.*, vol. 37, no. 6, pp. 543–557, 2005.
[4] X. Si, W. Wang, C. Hu, and D. Zhou, "Remaining useful life estimation-A review on the statistical data driven approaches," *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 1–14, Aug. 2011.
[5] Z. Chen, T. Xia, Y. Li, and E. Pan, "A hybrid prognostic method based on gated recurrent unit network and an adaptive Wiener process model considering measurement errors," *Mech. Syst. Signal Process.*, vol. 158, Feb. 2021, Art. no. 107785.
[6] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.
[7] L. Xiao, Z. Liu, Y. Zhang, Y. Zheng, and C. Cheng, "Degradation assessment of bearings with trend-reconstruct-based features selection and gated recurrent unit network," *Measurement*, vol. 165, Jun. 2020, Art. no. 108064.
[8] Y. Gao, Y. Wen, and J. Wu, "A neural network-based joint prognostic model for data fusion and remaining useful life prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 117–127, Jan. 2021.
[9] J. Deutsch and D. He, "Using deep learning-based approach to predict remaining useful life of rotating components," *IEEE Trans. Syst. Man Cybern.*, vol. 48, no. 1, pp. 11–20, Jan. 2018.

[10] H. Yang, K. Ding, R. C. Qiu, and T. Mi, "Remaining useful life prediction based on normalizing flow embedded sequence-to-sequence learning," *IEEE Trans. Rel.*, vol. 70, no. 4, pp. 1342–1354, Dec. 2021.

[11] S. Xiang, Y. Qin, J. Luo, H. Pu, and B. Tang, "Multicellular LSTM-based deep learning model for aero-engine remaining useful life prediction," *Rel. Eng. Syst. Saf.*, vol. 216, Jul. 2021, Art. no. 107927.

[12] H. Liu, Z. Liu, W. Jia, and X. Lin, "Remaining useful life prediction using a novel feature-attention-based end-to-end approach," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 1197–1207, Feb. 2021.

[13] M. Ma and Z. Mao, "Deep-convolution-based LSTM network for remaining useful life prediction," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1658–1667, Mar. 2021.

[14] R. Jin, Z. Chen, K. Wu, M. Wu, X. Li, and R. Yan, "Bi-LSTM-based two-stream network for machine remaining useful life prediction," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 3511110.

[15] K. Liu, N. Z. Gebraeel, and J. Shi, "A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 652–664, Jul. 2013.

[16] K. Liu and S. Huang, "Integration of data fusion methodology and degradation modeling process to improve prognostics," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 1, pp. 344–354, Jan. 2016.

[17] D. Wang, K. Liu, and X. Zhang, "A generic indirect deep learning approach for multisensor degradation modeling," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 3, pp. 1924–1940, Jul. 2022, doi: 10.1109/TASE.2021.3072363.

[18] H. Firpi and G. Vachtsevanos, "Genetically programmed-based artificial features extraction applied to fault detection," *Eng. Appl. Artif. Intell.*, vol. 21, no. 4, pp. 558–568, Aug. 2008.

[19] L. Guo, N. Li, F. Jia, and Y. Lei, "A recurrent neural network-based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, Feb. 2017.

[20] L. Chen, G. Xu, S. Zhang, W. Yan, and Q. Wu, "Health indicator construction of machinery based on end-to-end trainable convolution recurrent neural networks," *J. Manuf. Syst.*, vol. 54, pp. 1–11, Jan. 2020.

[21] J. Chen, H. Jing, Y. Chang, and Q. Liu, "Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process," *Rel. Eng. Syst. Saf.*, vol. 185, pp. 372–382, Jan. 2019.

[22] C. Song, K. Liu, and X. Zhang, "Integration of data-level fusion model and kernel methods for degradation modeling and prognostic analysis," *IEEE Trans. Rel.*, vol. 67, no. 2, pp. 640–650, Jun. 2018.

[23] K. Nguyen and K. Medjaher, "An automated health indicator construction methodology for prognostics based on multi-criteria optimization," *ISA Trans.*, vol. 113, pp. 81–96, Jul. 2021.

[24] L. Liao, "Discovering prognostic features using genetic programming in remaining useful life prediction," *IEEE Trans. Ind. Informat.*, vol. 61, no. 5, pp. 2464–2472, May 2014.

[25] K. Liu, A. Chehade, and C. Song, "Optimize the signal quality of the composite health index via data fusion for degradation modeling and prognostic analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 3, pp. 1504–1514, Jul. 2017.

[26] T. Yan, D. Wang, T. Xia, and L. Xi, "A generic framework for degradation modeling based on fusion of spectrum amplitudes," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 1, pp. 308–319, Jan. 2022.

[27] F. Wang, J. Du, Y. Zhao, T. Tang, and J. Shi, "A deep learning based data fusion method for degradation modeling and prognostics," *IEEE Trans. Rel.*, vol. 70, no. 2, pp. 775–789, Jun. 2021.

[28] J. Coble and J. W. Hines, "Identifying optimal prognostic parameters from data: A genetic algorithms approach," *Annu. Conf. PHM Soc.*, vol. 1, no. 1, pp. 1–11, 2009. [Online]. Available: http://papers.phmsociety.org/index.php/phmconf/article/view/1404

[29] B. Zhang, L. Zhang, and J. Xu, "Degradation feature selection for remaining useful life prediction of rolling element bearings," *Qual. Rel. Eng. Int.*, vol. 32, no. 2, pp. 547–554, Feb. 2016.

[30] Z. Chen, T. Xia, D. Zhou, and E. Pan, "A health index construction framework for prognostics based on feature fusion and constrained optimization," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 3523315.

[31] H. Lin, M. Tegmark, and D. Rolnick, "Why does deep and cheap learning work so well?," *J. Statist. Phys.*, vol. 168, no. 6, pp. 1223–1247, 2019.

[32] Z. Li, J. Wu, and X. Yue, "A shape-constrained neural data fusion network for health index construction and residual life prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 5022–5033, Nov. 2021.

[33] C. Song and K. Liu, "Statistical degradation modeling and prognostics of multiple sensor signals via data fusion: A composite health index approach," *IISE Trans.*, vol. 50, no. 10, pp. 853–867, May 2018.

[34] T. P. Nguyen et al., "Analysis and comparison of multiple features for fault detection and prognostic in ball bearings," in *Proc. PHM Soc. Eur. Conf.*, 2018, pp. 1–9.

[35] P. Baraldi P, G. Bonfanti, and E. Zio, "Differential evolution-based multi-objective optimization for the definition of a health indicator for fault diagnostics and prognostics," *Mech. Syst. Signal Process.*, vol. 102, pp. 382–400, Mar. 2018.

[36] Y. Shan, J. Liu, Y. Xu, and J. Zhou, "A combined multi-objective optimization model for degradation trend prediction of pumped storage unit," *Measurement*, vol. 169, Feb. 2021, Art. no. 108373.

[37] M. Ma, C. Sun, Z. Mao, and X. Chen, "Ensemble deep learning with multi-objective optimization for prognosis of rotating machinery," *ISA Trans.*, vol. 113, pp. 166–174, Jul. 2021.

[38] H. Mo and G. Iacca, "Multi-objective optimization of extreme learning machine for remaining useful life prediction," in *Proc. Int. Conf. Appl. Evol. Comput. (Part EvoStar)*, 2022, pp. 191–206.

[39] X. Zhang, L. Jin, C. Cui, and J. Sun, "A self-adaptive multi-objective dynamic differential evolution algorithm and its application in chemical engineering," *Appl. Softw. Comput.*, vol. 106, Jul. 2021, Art. no. 107317, doi: 10.1016/j.asoc.2021.107317.

[40] R. Fan, L. Wei, X. Li, J. Zhang, and Z. Fan, "Self-adaptive weight vector adjustment strategy for decomposition-based multi-objective differential evolution algorithm," *Soft Comput.*, vol. 24, no. 17, pp. 13179–13195, Feb. 2020.

[41] Y. Xu, D. Pi, S. Yang, Y. Chen, S. Qin, and E. Zio, "An angle-based bi-objective optimization algorithm for redundancy allocation in presence of interval uncertainty," *IEEE Trans. Autom. Sci. Eng.*, to be published, doi: 10.1109/TASE.2022.3148459.

[42] Q. Dong, L. Cui, and S. Si, "Reliability and availability analysis of stochastic degradation systems based on bivariate Wiener processes," *Appl. Math. Model.*, vol. 79, pp. 414–433, Oct. 2020.

[43] J. Hu, Q. Sun, Z. S. Ye, and X. Ling, "Sequential degradation-based burn-in test with multiple periodic inspections," *Front. Eng. Manage.*, vol. 8, no. 4, pp. 519–530, Jun. 2021.

[44] Y. Zhang, Y. Yang, H. Li, X. Xiu, and W. Liu, "A data-driven modeling method for stochastic nonlinear degradation process with application to RUL estimation," *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 52, no. 6, pp. 3847–3858, Jun. 2022, doi: 10.1109/TSMC.2021.3073052.

[45] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, and N. Zerhouni, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," presented at the *IEEE Int. Conf. Prognostics Health Manage.*, Denver, CO, USA, 2012.

[46] D. Lyu, G. Niu, E. Liu, T. Yang, G. Chen, and B. Zhang, "Uncertainty management and differential model decomposition for fault diagnosis and prognosis," *IEEE Trans. Ind. Electron.*, vol. 69, no. 5, pp. 5235–5246, May 2022.

[47] D. Liu, J. Zhou, H. Liao, Y. Peng, and X. Peng, "A health indicator extraction and optimization framework for lithium-ion battery degradation modeling and prognostics," *IEEE Trans. Syst. Man Cybern.-Syst.*, vol. 45, no. 6, pp. 915–928, Jun. 2015.

[48] IEEE PHM, "Prognostic challenge-outlines, experiments, scoring of results, winners," 2012, Accessed: Sep. 9, 2016. [Online]. Available: http://www.femto-st.fr/f/d/IEEEPHM2012-Challenge-Details.pdf

[49] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Jun. 2001.

[50] 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.

Zhen Chen (Member, IEEE) received the B.S. degree in quality and reliability engineering from Beihang University, Beijing, China, in 2014 and the Ph.D. degree in mechanical engineering (industrial engineering) from Shanghai Jiao Tong University, Shanghai, China, in 2019.

He is currently an Assistant Professor with the Department of Industrial Engineering and Management, Shanghai Jiao Tong University. He was a Visiting Scholar with the Department of Energy, Politecnico di Milano, Milan, Italy. His research interests include system reliability assessment, degradation modeling, and prognostics and health management.

Di Zhou received the B.S. and Ph.D. degrees in mechanical engineering from Northeastern University, Shenyang, China, in 2012 and 2017, respectively.

He was a Postdoctoral Research with the Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai, China. He is currently a Lecturer with College of Mechanical Engineering, Donghua University, China. His research interests include mechanical reliability analysis and prognostics and health management.

Enrico Zio (Senior Member, IEEE) received the B.S. and Ph.D. degrees in nuclear engineering from the Politecnico di Milano, Milan, Italy, in 1991 and 1995, respectively, the M.S. degree in mechanical engineering from the University of California at Los Angeles (UCLA), Los Angeles, CA, USA, in 1995, and the Ph.D. degree in nuclear engineering from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1998.

He is currently a Full Professor with the Centre for Research on Risk and Crises, École des Mines, ParisTech, PSL University, Paris, France, a Full Professor and the President of the Alumni Association, Politecnico di Milano, and an Eminent Scholar with Kyung Hee University, Seoul, South Korea. His research interests include modeling of the failure—repair—maintenance behavior of components, complex systems for the analysis of their reliability, maintainability, safety, vulnerability, resilience, and security characteristics, and the development and use of Monte Carlo simulation methods, artificial techniques, and optimization heuristics.

Tangbin Xia (Member, IEEE) received the Ph.D. degree in mechanical engineering (industrial engineering) from Shanghai Jiao Tong University, Shanghai, China, in 2014.

He was a Visiting Scholar with the Georgia Institute of Technology, Atlanta, GA, USA, and the University of Michigan, Ann Arbor, MI, USA. He is currently the Deputy Director of the Department of Industrial Engineering and an Associate Professor with the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai, China. His research interests include advanced manufacturing, prognostics and health management (PHM), and intelligent maintenance systems.

Ershun Pan received the B.S. and M.S. degrees in mechanical design and manufacturing from Northeastern University, Shenyang, China, in 1997 and the Ph.D. degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2000.

He is currently a Professor and the Header with the Department of Industrial Engineering and Management, Shanghai Jiao Tong University. His current research interests include quality control, reliability engineering and predictive maintenance, production system planning and design.