Asia-Pacific Journal of Operational Research

A Semi-supervised Nonlinear Approach for Health Assessment with Prognostic Accuracy Integration

Zihao Zhao
School of Management, Beijing institute of Technology
Beijing, 100081,China
zihaozhao@bit.edu.cn

Zhen Li
China Mobile Information Technology Center
Beijing, 102200, China
zhen.li@pku.edu.cn

Junbo Son
Lerner College of Business & Economics, University of Delaware
Delaware, Newark, 19716, USA
junboson@udel.edu

Jianguo Wu*
College of Engineering, Peking University
Beijing, 100871,China
j.wu@pku.edu.cn

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)
Published (Day Month Year)

Sensor networks are crucial in engineering for monitoring conditions and predicting failures. However, conventional signal fusion techniques for constructing health indices (HIs) typically depend on linear combinations and predefined parametric models, rendering them inadequate for capturing the intricate nonlinear relationships between health conditions and sensor data. Furthermore, the prediction of remaining useful life (RUL)—a critical task intimately linked to the underlying health condition—is often neglected during the design phase of HI development. To address these limitations, this study introduces a novel semi-supervised nonlinear data fusion framework that enhances both health assessment and RUL prediction accuracy. Our approach leverages a semi-supervised neural network to model the complex nonlinear dependencies between diverse sensor signals and health conditions. A distinctive feature of this framework is the integration of the variance of both predicted and observed HIs at the time of failure into the loss function, ensuring consistency and reliability in predictions. Additionally, we incorporate regularization terms informed by engineering domain knowledge to enhance the robustness of

*Corresponding Author

the health assessment model. A comprehensive evaluation using the NASA C-MAPSS dataset demonstrates the superior performance of the proposed method compared to existing approaches, underscoring its potential for advancing prognostic accuracy in engineering systems.

Keywords: Semi-supervised neural network; data fusion; health assessment; remaining useful life prediction

1. Introduction

The health condition of engineering systems and mechanical components inevitably degrades over time, even under normal operating conditions. Although this degradation may not be immediately visible, it accumulates until the system or component ceases to function. This failure results in costly production downtime, significant financial losses, and potential safety hazards. Therefore, rigorous health assessment and accurate prediction of remaining useful life (RUL) are vital for enabling timely preventive actions to avoid catastrophic failures (Liu *et al.*, 2013; Taşcı *et al.*, 2023; Jain and Priya, 2005; Liu *et al.*, 2019b; Cai *et al.*, 2021).

Due to their importance, degradation modeling and RUL prediction have been extensively researched over the past few decades within the field of condition-based maintenance. Early studies predominantly relied on data from a single sensor, assuming it could adequately represent the true health status of the system (Nelson, 1990; Wei *et al.*, 2021; Wang *et al.*, 2024b). However, as modern engineering systems have become increasingly complex, it has become clear that a single sensor's limited data cannot fully capture the entire degradation process (Ramasso and Saxena, 2014; Yuan *et al.*, 2025). With advancements in information and sensing technologies, the use of multiple sensors has become standard, allowing for collaborative monitoring of in-service units' health from diverse physical perspectives.

To take the full advantage of such multi-sensor monitoring system and to implement a health condition assessment, a proper data (signal) fusion is essential. A direct outcome of the data fusion is a composite indicator, namely, the health index (HI) (Liu *et al.*, 2013; Franklin, 2005). The HI intuitively visualizes the underlying health condition, keeping the decision-makers better informed of the degradation status of the engineering system or the in-service unit. In the recent past, various data fusion approaches have been proposed specifically for the health assessment and the HI construction. Zhang *et al.* (Zhang *et al.*, 2019) constructed an HI based on the Mahalanobis distance to measure how far away the multi-sensor data is deviating from the normal state. Instead of the Mahalanobis distance, Yu *et al.* (Yu and Wang, 2009) used the Minimum Quantization Error to evaluate the amount of deviation from the normal state in their HI construction process. In common, these methods measure the distance between the current monitoring signal data with the ones in the normal health state, and the failure is declared when the signal exceeds a pre-defined threshold. Despite the promising results reported in the literature, prior studies often suffer from a few limitations. These methods are unsupervised and are more or less subjective, in that both the deviation metric and the failure thresh-

old are chosen arbitrarily. It is not straightforward to determine the best deviation measure for a given engineering application because, to the best of our knowledge, there is no systematic procedure for selecting the optimal distance measure. Therefore, it may not be easy to implement these methods in practice due to the lack of necessary justifications.

In fact, failure is typically a direct outcome of the degradation process; hence, coupling the HI and the RUL or failure information together is a natural modeling approach given their close relationship. It is worth noting that in several supervised RUL prediction approaches, the RUL or alternatively the percentage of remaining life is treated as the HI itself (Guo *et al.*, 2017; Zhang *et al.*, 2018; Chen *et al.*, 2020). Although they are highly correlated, however, the HI and RUL are two very distinctive health assessment metrics. For example, a working unit typically stays healthy with only minimal degradation in the early stage of its life cycle but deteriorates rapidly as it gets closer to the end of its life. This highly nonlinear degradation pathway is not applicable to the RUL because the RUL of a working unit decreases linearly until it reaches zero at the time of failure. Therefore, the HI defined in this paper is different from the one defined by RUL. To this end, some researches incorporated the actual failure information in the HI development by considering that the HIs should be consistent at the failure time for all units under the same failure mode/operational condition. For instance, Liu *et al.* (Liu *et al.*, 2013) constructed the HI through a linear combination of multiple sensor signals, where the weights for each signal are optimized by jointly minimizing the HI variance at the failure time while maximizing the degradation monotonicity. Later, Song *et al.* (Song *et al.*, 2017) presented a kernel-based method to capture the nonlinear relationship between sensor signals and the health condition. However, all these methods assume that the underlying HI has a parametric form, which is too restrictive. Having a parametric form makes analytical studies a lot more convenient but such assumptions often make the method impractical in many modern engineering applications. For example, Wang *et al.* (Wang *et al.*, 2024a) proposed a flexible degradation modeling approach via features extraction that characterize degradation progression and construct local models to describe the relationship between these features and failure time. Moreover, since the RUL prediction is a crucial subsequent task in the preventive maintenance, it is desirable to consider the prognostic accuracy in the HI construction. The aforementioned approaches only guarantee the consistency of the observed HIs at the failure time, but cannot ensure the predicted HIs to be consistent when failure occurs. In this paper, we differentiate the observed and predicted HIs. The term "observed HI" refers to the calculated HI using the *observed* sensing signals, while the "predicted HI" indicates the estimated HI based on the *predicted* sensing signals. Technically, the HI used in the RUL prediction must be the predicted HI, because the sensing signal data at a future time instant can never be observed at the time of prediction. Therefore, the errors introduced in the HI prediction stage carry over to the later RUL prediction stage. Due to this nature, it is important to keep the RUL prediction in mind at

the design stage of the HI construction method. However, existing methods in the literature rarely address this issue.

To fill the above mentioned research gap in the current literature, we propose a semi-supervised nonlinear data fusion approach for health assessment with the prognostic accuracy of the RUL considered. Specifically, a neural network is established to model the complex nonlinear relationship between multi-sensor signals and underlying health condition. For the loss function, we propose to use the variance of both the predicted and the observed HIs at the time of failure as determinant terms, while other domain knowledge regarding the degradation properties, i.e., monotonicity and degradation acceleration are applied as two regularizers. The proposed loss function not only improves the quality of the established HI but also provides more satisfactory prognostic accuracy in the subsequent RUL prediction.

The rest of this paper is organized as follows. Section 2 gives the problem description of the health assessment and the HI construction introducing the research setting and background. The technical details of our health assessment framework are provided in Section 3. Using a widely-accepted C-MAPSS data from the National Aeronautics and Space Administration (NASA), in Section 4, we demonstrate the advantageous features of our proposed method through a comprehensive case study, and the performance evaluation results are reported as well along with in-depth discussions. Lastly, Section 5 concludes the paper with a brief discussion on the limitations of our method and some of the potential future research directions.

2. Problem Description

In this section, we first give the mathematical formulation of this problem, and then introduce the formal definition and assumption of HI.

Let $x_{n,t} = (x_{n,t,1}, \cdots, x_{n,t,S})$ denote the vector of all S sensor measurement data for unit n at time t. Specifically, $x_{n,t,s}$ denotes the observation for unit n, collected from sensor s at time t. Then the multi-sensor data $\mathbf{X}_n$ for unit n is given as

$$\mathbf{X}_n = (\boldsymbol{x}_{n,1}, \boldsymbol{x}_{n,2}, \ldots, \boldsymbol{x}_{n,T_N})'. \tag{1}$$

The total number of units in the historical dataset is N, and $\boldsymbol{T} = (T_1, \ldots, T_N)$ represents the total lifetimes (failure time) of these N historical units. The corresponding multi-sensor data is denoted by $\boldsymbol{D} = \{\mathbf{X}_n, n = 1, \ldots, N\}$.

Based on the multi-channel signals, a composite HI can be developed. To capture the complex nonlinear relationship, we propose a flexible nonlinear data fusion method $f(\cdot)$ to integrate the multi-sensor signals for unit n into a composite index. The constructed HI captures the underlying health condition and it is defined as:

$$h(\boldsymbol{x}_{n,t}) = f(x_{n,t,1}, x_{n,t,2}, \cdots, x_{n,t,S}),$$

where $h(\boldsymbol{x}_{n,t})$ represents the HI of unit n at time t.

The HI construction is basically a dimensionality reduction process, which maps the high-dimensional sensor data to a one-dimensional index that satisfies a set of

desirable properties. Fig. 1 illustrates the HI construction process based on three sensor signals. In the original space, one or multiple failure hypersurfaces exist. Once the high-dimensional degradation signals cross one or more hypersurfaces and enter the failure zone, the working unit fails. By constructing the HI, the degradation path is reduced to a one-dimensional index, and the failure hypersurfaces are reduced to a single failure threshold. From this perspective, the RUL prediction can be viewed as the integration of time series prediction task and classification task, i.e., predicting the future degradation trajectory and judging whether it is in the failure zone.

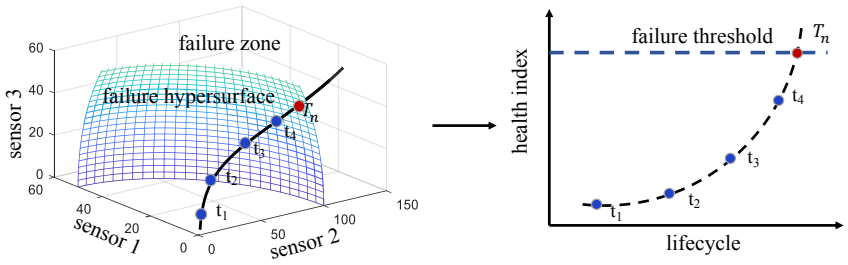

Fig. 1. Illustration of HI Construction as a dimensionality reduction process.

3. Methodology

We aim to provide a practical solution for the real-world health assessment problem. To achieve the goal, we make sure that the method proposed in this paper have a set of properties that are highly desirable in diverse engineering practices. In this section, we first discuss those properties of the HI. Then, our semi-supervised neural network-based health assessment framework is defined. The loss function used for the model learning is designed in a way that realizes the key properties. For the parameter estimation, we establish the tailored Adam algorithm for our model specifically. For the RUL prediction, we first need to predict the future sensor signals so that the HI at the future time instant can be predicted accordingly. We adopt the linear mixed effects model with Bayesian updating for the signal prediction. The overall framework of the proposed method is provided in Figure 2 and the methodological details are provided in the following subsections.

3.1. *Key Properties for HI Construction*

The purpose of constructing the HI is mainly for establishing a quantitative metric that correctly reflects the true underlying health condition that is typically unknown and unobservable in practice. To make the constructed HI analytically tractable and to ensure reasonable performance, there are a few properties that are shared by many existing HIs in the literature. In this subsection, we describe those popular

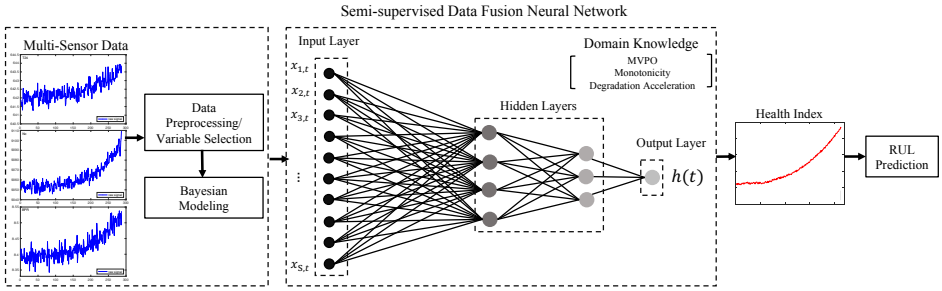

Fig. 2. Overall framework of the proposed methodology.

properties and highlight the additional useful properties that our method uniquely offers.

When it comes to the health assessment and failure prediction, it is common to assume that all units have a similar level of health condition at the time of failure (Liu *et al.*, 2013; Song *et al.*, 2017). Although the degradation paths may be considerably different due to the existence of heterogeneity among units, the health condition of all units at failure time should be close to a specific value, namely the failure threshold. Based on this rationale, many HIs available in the literature commonly have Property 1:

Property 1 *(Minimum variance of the HIs at the failure time). The variance of the HIs for all units at the failure time should be minimum under the same operational condition and failure mode.*

Thanks to Property 1, the failure prediction task can be formulated as a typical binary classification algorithm where the decision rule is the constructed HI and the decision boundary is the failure threshold as previously shown in Fig. 1. However, Property 1 only guarantees the consistency of the observed HI at the failure time. As discussed earlier, the RUL prediction depends heavily on the predicted HI rather than the observed HI. Because Property 1 cannot ensure the same level of consistency for the predicted HI at future time instances, satisfactory accuracy of RUL prediction may not be achievable based solely on Property 1. To this end, we argue that the HI construction process must incorporate a new and improved property (Property 1*) considering both the observed and predicted HIs:

Property 1* *(Minimum variance of the predicted and observed HIs at the failure time). Under the same operational condition and failure mode, the variance of both the observed and predicted HIs at the failure time should be minimum.*

Property 1* ensures the minimum variance of the predicted and observed HIs (MVPO) at the time of failure. We wanted our HI construction method to have the MVPO property so that the subsequent RUL prediction can be done with satisfactory accuracy.

Besides the minimal variance property of HI, the monotonicity is another property on the degradation process that is frequently used in the existing methods, which is based on the fact that the degradation process is often irreversible in nature. The monotonicity property is given in Property 2 as follows:

Property 2 *(Monotonicity). The HI is monotonic within the entire degradation process.*

This monotonicity property of the HI is highly desirable because, in many cases, the original sensor signals are often contaminated by non-monotonic measurement noises. The noisy sensor signals make it difficult for us to identify a meaningful trend and that is why being able to visualize the degradation path through the constructed HI is helpful in practice. Therefore, to make the constructed HI show the degradation trend clearly with minimal fluctuations, the existing HIs in the literature typically offer the monotonicity property (Song *et al.*, 2017; Song and Liu, 2018; Liu *et al.*, 2013).

Property 2 is significantly useful in practice. However, we would like to emphasize that our method offers a more desirable property than the conventional monotonicity property. For many engineering systems, degradation often shows an accelerated trend throughout the life cycle (Wen *et al.*, 2018a, 2019; Zhou *et al.*, 2014). To correctly reflect this physical mechanism of degradation process, the absolute value of the slope for the constructed HI curve should increase over time. In this spirit, we propose Property 2*:

Property 2* *(Monotonic and accelerated degradation). The HI is monotonic within the entire degradation process, and the degradation rate should increase as the operating unit approaches to the end of its life cycle.*

Developing a new HI that has Property 2* can potentially improve the performance of RUL prediction. Property 2* can enlarge the range of HI during the entire degradation process due to the accelerated rate, which yields a higher signal-to-noise ratio (SNR) (Song *et al.*, 2017). The high SNR can help with improving the accuracy of the RUL prediction especially for noise-contaminated data as shown later in the case study.

3.2. *Semi-supervised Neural Network Model*

We propose a semi-supervised neural network model for the HI construction with the two key properties (MVOP and monotonic accelerated degradation) in mind. With the collected multiple sensor signals $x_{n,t,1}, x_{n,t,2}, \cdots, x_{n,t,S}$, the health condition for a unit n at time t is characterized by the HI $h\left(\boldsymbol{x}_{n,t}\right)$. Considering that the underlying degradation mechanism could be very complicated, it could be challenging to use an analytical expression to correctly describe the relationship between the multi-sensor signals and the HI. Thus, we utilize the artificial neural network as the fusion function $f(\cdot)$, which is capable of approximating any nonlinear functions (Chen and Chen, 1995). Unlike the traditional neural network-based prognostics methods that give only the predicted RUL for each unit as an outcome (Guo *et al.*, 2017; Zhang *et al.*, 2018; Chen *et al.*, 2020), the proposed neural network can generate the entire health condition profile for each unit.

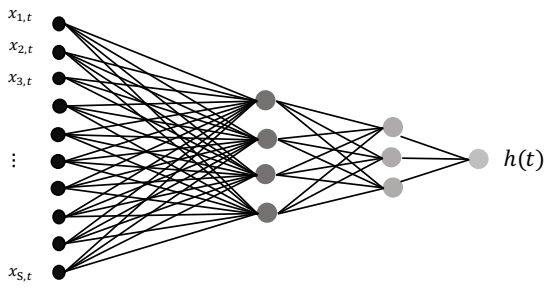

Fig. 3. Illustration of the neural network structure for HI construction.

Fig. 3 illustrates an example of the proposed semi-supervised neural data fusion model. The model input is a collection of the observed multi-sensor signals, while the model output is a composite HI. For the activation functions, the hyperbolic tangent function $\tanh\left(y\right) = \frac{e^{y}-e^{-y}}{e^{y}+e^{-y}}$ is employed between the input layer and the middle layers (Sibi *et al.*, 2013). Between the last hidden layer and the model output, the identity activation function is applied to generate the HI.

The conventional supervised learning approaches require the HI values to perform the health assessment. However, the HI are not available in practice, making the supervised learning approaches not directly applicable. Therefore, we propose a novel semi-supervised neural network model. The key properties (Property 1* and Property 2*) are attained in our model. First, when designing our loss function, we incorporate the variances of both the predicted and the observed HIs at the failure time as the major losses. Then, Property 2* (monotonic accelerated degradation) is obtained by the two regularization terms in the loss function.

Suppose we have

$$\boldsymbol{x}^*_{n,t} = \begin{cases} \hat{\boldsymbol{x}}^{1:t}_{n,T_n}, 1 \leq t < T_n, \\ \boldsymbol{x}_{n,T_n}, t = T_n, \end{cases} \tag{2}$$

where $\hat{\boldsymbol{x}}^{1:t}_{n,T_n}$ denotes the predicted sensor signals at the failure time based on the historical observations from time 1 to t for unit n. Then, Property 1* (MVPO) can be achieved by:

$$\sum_{n=1}^{N}\sum_{t=1}^{T_n}[h(\boldsymbol{x}^*_{n,t}) - \overline{h}]^2, \tag{3}$$

where $\bar{h} = \frac{1}{\sum_{n=1}^{N} T_n}\sum_{t=1}^{T_n}\sum_{n=1}^{N} h\left(\boldsymbol{x}^*_{n,t}\right)$. Note that when minimizing Eq. (3), we can set $\overline{h}$ to be an arbitrary value. Suppose $h\left(\boldsymbol{x}^*_{n,t}\right) = c_0 + \sum_{i=1}^{K} c_i w_{n,i}$, where $w_{n,i}$ is the ith node of the last hidden layer for unit n; c_i and c_0 are the weight and bias term respectively. Then, it is easy to see that minimizing the variance term in Eq. (3) does not involve c_0 because both $h(\boldsymbol{x}^*_{n,t})$ and $\overline{h}$ have c_0. Therefore, we can arbitrarily select a value for c_0. Here we set $\overline{h} = 1$ to determine c_0, so that we can conveniently set the failure threshold to be 1. In this setting, Eq. (3) can be simplified to

$$\sum_{n=1}^{N}\sum_{t=1}^{T_n}[h(\boldsymbol{x}^*_{n,t}) - 1]^2. \tag{4}$$

This loss function naturally makes the entire HI construction task a semi-supervised learning problem because, although the true health conditions before failure are unavailable, they can be regarded as labeled at the time of failure. Our loss function includes not only the observed HI but also the predicted HI at the time of failure. Therefore, the loss function we propose in this paper is a generalization of the traditional loss function. If all the sensor signals can be perfectly predicted, and all units have the same lifetime, our loss function will be simplified to the traditional loss function. The conventional loss function gives only Property 1 whereas our generalized loss function attains Property 1* which is the key for improving the accuracy of the RUL prediction.

To further improve the quality of HI construction, we use exponentially weighted loss function expressed as

$$\sum_{n=1}^{N}\sum_{t=1}^{T_n}\exp\left[k\left(t - T_n\right)\right]\left[h\left(\boldsymbol{x}^*_{n,t}\right) - 1\right]^2, \tag{5}$$

where k is a tuning parameter controlling the decaying speed or half-life. The rationale of using the exponential weight is in two folds. First, the lifetimes vary across different units. Due to this nature, it is possible that the loss in Eq. (4) is may be dominated by a few units with long lifetimes. Using exponentially weighted loss could help mitigate the issue. Second, as the unit approaches to its failure time, the

prognostic accuracy becomes increasingly more important. Therefore, it is beneficial to put higher weight on the loss closer to the failure times (Liu *et al.*, 2013).

As mentioned, Property 2* is ensured by two regularization terms in the loss function. The monotonicity property is embedded in the following regularization (penalty) term

$$\sum_{n=1}^{N}\sum_{t=2}^{T_n}\frac{1}{T_n-1}\left[h\left(\boldsymbol{x}_{n,t-1}\right)-h\left(\boldsymbol{x}_{n,t}\right)\right]_+, \tag{6}$$

where $[x]_+ = \max(0,x)$. If $h\left(\boldsymbol{x}_{n,t-1}\right) < h\left(\boldsymbol{x}_{n,t}\right)$, then $\left[h\left(\boldsymbol{x}_{n,t-1}\right)-h\left(\boldsymbol{x}_{n,t}\right)\right]_+ = 0$, i.e., there is no penalty. On the other hand, if $h\left(\boldsymbol{x}_{n,t-1}\right) > h\left(\boldsymbol{x}_{n,t}\right)$, i.e., the monotonicity is violated, this loss term linearly increases by the amount of $h\left(\boldsymbol{x}_{n,t-1}\right)-h(\boldsymbol{x}_{n,t})$. Because the total lifetime for each unit could vary significantly across different units, we divide the penalty amount by T_n-1 for each unit to reduce the unbalance caused by the lifetime difference.

Similarly, to attain the degradation acceleration property, the following penalty term is applied in the loss function,

$$\sum_{n=1}^{N}\sum_{t=3}^{T_n}\frac{1}{T_n-2}\left[\left(h(\boldsymbol{x}_{n,t-1})-h(\boldsymbol{x}_{n,t-2})\right)-\left(h(\boldsymbol{x}_{n,t})-h(\boldsymbol{x}_{n,t-1})\right)\right]_+. \tag{7}$$

The overall loss function is then formulated by integrating the three terms in Eq. (4)-(7) as:

$$\begin{aligned} L\left(f,\boldsymbol{D},\boldsymbol{\theta}\right) = &\sum_{n=1}^{N}\sum_{t=1}^{T_n}\exp\left[k\left(t-T_n\right)\right]\left[h(\boldsymbol{x}^*_{n,t})-1\right]^2 \\ &+\lambda_1\sum_{n=1}^{N}\sum_{t=2}^{T_n}\frac{1}{T_n-1}\left[h\left(\boldsymbol{x}_{n,t-1}\right)-h\left(\boldsymbol{x}_{n,t}\right)\right]_+ \\ &+\lambda_2\sum_{n=1}^{N}\sum_{t=3}^{T_n}\frac{1}{T_n-2}\left[\left(h(\boldsymbol{x}_{n,t-1})-h(\boldsymbol{x}_{n,t-2})\right)\right. \\ &\left.-\left(h(\boldsymbol{x}_{n,t})-h(\boldsymbol{x}_{n,t-1})\right)\right]_+, \end{aligned} \tag{8}$$

where λ_1 and λ_2 are the penalty coefficients or weights to balance the three terms.

3.3. *Model Training through Customized Adam Algorithm*

The objective function is optimized through the gradient descent with back propagation algorithm. The stochastic gradient descent (SGD) algorithm is adopted to minimize the loss function. The SGD algorithm randomly extracts a subset of training samples to compute the gradient in each iteration for reducing the computational burden (Bottou, 2010). In the SGD algorithm, one of the critical steps is to calculate the partial derivatives of the loss function. Since the loss function in

Eq. (8) is not differentiable at certain points, to apply the SGD algorithm, we use the sub-gradient at those non-differentiable points. The gradient or sub-gradient of the loss function for unit n can be derived as

$$\begin{aligned}\frac{\partial L\left(f,\boldsymbol{D},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}} &= 2\sum_{t=1}^{T_n}\exp\left[k\left(t-T_n\right)\right]\left[h\left(\boldsymbol{x}_{n,t}^{*}\right)-1\right]\frac{\partial h\left(\boldsymbol{x}_{n,t}^{*}\right)}{\partial\boldsymbol{\theta}}\\ &+\frac{\lambda_1}{T_n-1}\sum_{t=2}^{T_n}\delta\left[h\left(\boldsymbol{x}_{n,t-1}\right)-h\left(\boldsymbol{x}_{n,t}\right)\right]\times\frac{\partial\left(h\left(\boldsymbol{x}_{n,t-1}\right)-h\left(\boldsymbol{x}_{n,t}\right)\right)}{\partial\boldsymbol{\theta}}\\ &+\frac{\lambda_2}{T_n-2}\sum_{t=3}^{T_n}\delta\left[\left(h(\boldsymbol{x}_{n,t-1})-h(\boldsymbol{x}_{n,t-2})\right)-\left(h(\boldsymbol{x}_{n,t})-h(\boldsymbol{x}_{n,t-1})\right)\right]\\ &\times\frac{\partial\left[\left(h(\boldsymbol{x}_{n,t-1})-h(\boldsymbol{x}_{n,t-2})\right)-\left(h(\boldsymbol{x}_{n,t})-h(\boldsymbol{x}_{n,t-1})\right)\right]}{\partial\boldsymbol{\theta}},\end{aligned}\tag{9}$$

where $\delta(x)$ is the indicator function given by

$$\delta\left(x\right)=\begin{cases}1, & x>0\\ 0, & x\leq 0\end{cases}$$

During the iterative process, the gradient or sub-gradient of the loss function is computed by the back-propagation algorithm. To improve the computational efficiency with less memory requirement and to make the algorithm easy to be implemented in practice, we establish an optimization algorithm tailored to our method for efficient parameter estimation. Our algorithm is based on the adaptive moment estimation optimizer (Adam). The Adam, which can be viewed as an extended version of the SGD algorithm (Kingma and Ba, 2014), is a widely used stochastic optimization algorithm based on adaptive estimation of low-order moments and has been extensively used for model training in deep learning applications (Miao *et al.*, 2019; Liu *et al.*, 2019a). The tailored Adam algorithm for our method is presented in Algorithm 1, in which $\boldsymbol{q}_k^2$ represents the element-wise square.

Algorithm 3.1 Engineering-informed Neural Data Fusion Model Fitting Using Tailored Adam.

$\eta = 0.001$, $\gamma_1 = 0.9, \gamma_2 = 0.999$ and $\epsilon = 10^{-8}$.**Input**: training samples $\boldsymbol{D}$, default ADAM algorithm parameters: step size, η; time step, $k \leftarrow 0$; exponential decay rates for the two moment estimates, γ_1, γ_2; the 1st and 2nd moment vector, $\boldsymbol{m}_0 \leftarrow \boldsymbol{0}$, $\boldsymbol{v}_0 \leftarrow \boldsymbol{0}$; the initial value for $\boldsymbol{\theta}$ to be optimized, $\boldsymbol{\theta}_0$.**Output**: optimized Model Parameters $\boldsymbol{\theta}_k$. $N \neq 0$ Calculate the gradient or sub-gradient $\frac{\partial L(f,\mathbf{X}_n,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}$ based on Eq. (9) $\boldsymbol{m}_k \leftarrow \gamma_1 \cdot \boldsymbol{m}_{k-1} + (1-\gamma_1)\cdot \boldsymbol{q}_k$(update biased 1st moment estimate) $\boldsymbol{v}_k \leftarrow \gamma_2 \cdot \boldsymbol{v}_{k-1} + (1-\gamma_2)\cdot \boldsymbol{q}_k^2$ (update biased 2nd moment estimate) $\widehat{\boldsymbol{m}}_k \leftarrow \frac{\boldsymbol{m}_k}{1-\gamma_1^k}$ (compute bias-corrected 1st moment estimate) $\widehat{\boldsymbol{v}}_k \leftarrow \frac{\boldsymbol{v}_k}{1-\gamma_2^k}$ (compute bias-corrected 2nd raw moment estimate) $\boldsymbol{\theta}_k \leftarrow \boldsymbol{\theta}_{k-1} - \eta\cdot\widehat{\boldsymbol{m}}_k/(\sqrt{\widehat{\boldsymbol{v}}_k}+\epsilon)$ (update parameters)**Return** $\boldsymbol{\theta}_k$

3.4. *Degradation Modeling and Prediction*

The RUL prediction is based on the predicted HI, and the HI is constructed by integrating multiple sensor signals. Therefore, it is critical to have a model for accurately predict the future sensor signals considering the unique characteristics of each individual unit. In this subsection, to complete our health assessment framework, we introduce a stochastic degradation modeling method based on a Bayesian updating approach which updates its model parameters based on the sensor signals collected from a specific unit of our interest. The historical sensor signals are modeled by the linear mixed effect model at the offline stage and then the model parameters are updated through the Bayesian method at the online stage. In this way, the model captures both the population-level general characteristics of degradation process and the heterogeneity among different units. Due to the simplicity and flexibility, the polynomial functions such as quadratic form are widely applied for modeling the degradation path (Liu *et al.*, 2013; Song *et al.*, 2017; Gao *et al.*, 2020). Therefore, without loss of the overall framework, here we use a linear model of a polynomial form to capture the sensor signal. In practice, the parametric form of the degradation path can be determined by the users for the specific engineering application. For example, if the linear model does not work, we can try other methods, such as LSTM (Zhang *et al.*, 2018), time series based prediction (Bontempi *et al.*, 2013), Gaussian process regression (Hong *et al.*, 2015).

Then, in the offline modeling stage, the linear mixed effects model with a polynomial structure for one unit is defined as

$$x_t = \sum_{k=0}^{q} \beta_k t^k + \varepsilon_t, \tag{10}$$

where q is the order of the polynomial model, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_q)'$ is the corresponding regression parameter, and ε_t is a normally distributed noise term with mean 0 and variance σ^2. To capture the noise heterogeneity, we can also assume σ^2 to be random. Here we assume $\sigma^2 \sim \text{IG}(a_1, a_2)$, where IG denotes the inverse Gamma distribution. $\boldsymbol{\beta}$ is specified as $\boldsymbol{\beta}|\sigma^2 \sim N(\boldsymbol{\mu}_0, \sigma^2\boldsymbol{\Sigma}_0)$. The model parameters $(a_1, a_2, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ are computed by the two-step estimation method (Wen *et al.*, 2018b), where the first step is to fit the sensor signals independently through the maximum likelihood estimation (MLE) method, and the second step is to estimate the hyperparameters based on the fitted parameters.

With the estimated hyperparameters obtained from the historical data at the offline stage, the individual model parameters can be updated according to the newly acquired sensor signals from a specific individual unit. First, we define

$$\boldsymbol{\Psi}_{1,t} = \begin{bmatrix} 1 & 1 & \cdots & 1^q \\ 1 & 2 & \cdots & 2^q \\ \cdots & \cdots & \cdots & \cdots \\ 1 & t & \cdots & t^q \end{bmatrix}$$

Suppose the sensor signals from time 1 to t is $y_{1:t} = (y_1, y_2, \ldots, y_t)$, then we get

$$\sigma^2 \mid y_{1:t} \sim \mathrm{IG}\left(a_1 + \tfrac{t}{2}, a_2 + \tfrac{H_t}{2}\right),$$
$$\left(\boldsymbol{\beta} \mid \sigma^2, y_{1:t}\right) \sim N(\boldsymbol{\mu}_t, \sigma^2 \boldsymbol{\Sigma}_t^{\mathrm{T}}),$$

where

$$\boldsymbol{\Sigma}_t = \left(\boldsymbol{\Psi}_{1,t}^{\mathrm{T}} \boldsymbol{\Psi}_{1,t} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1},$$

$$\mathbf{M}_t = \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 + \boldsymbol{\Psi}_{1,t} y_{1:t},$$

$$\boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t \mathbf{M}_t,$$

$$H_t = \|y_{1:t}\|^2 + \boldsymbol{\mu}_0^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 - \mathbf{M}_t^{\mathrm{T}} \boldsymbol{\Sigma}_t \mathbf{M}_t.$$

The proof of Eq. (3.4) can be found in the Appendix A. Based on the updated model, we obtained the predictive distribution of future observations $y_{t+1:t+L}$ given the historical observations up to the time t as:

$$y_{t+1:t+L} \mid y_{1:t} \sim MT\left(2a_1 + t, \boldsymbol{\Psi}_{t+1,t+L} \boldsymbol{\mu}_t, \quad \frac{2a_2 + H_t}{2a_1 + t}(\boldsymbol{I} + \boldsymbol{\Psi}_{t+1,t+L} \boldsymbol{\Sigma}_t \boldsymbol{\Psi}_{t+1,t+L}^{\mathrm{T}})\right),$$

where MT denotes the multivariate t-distribution. The three parameters in the multivariate t-distribution are the degree of freedom, mean, and shape matrix. Mathematical details are provided in Appendix B. With this predictive distribution, we can effectively predict the evolution of the degradation path and use it for the RUL prediction.

4. Case Study

In this section, we demonstrate the proposed method and evaluate its performance using a well-known C-MAPSS dataset from the NASA.

4.1. *Overview of the C-MAPSS Dataset*

C-MAPSS is a simulation tool developed by the NASA for high-fidelity simulation of the large commercial aircraft engine, whose health condition is continuously monitored by multiple embedded sensors (Saxena *et al.*, 2008b). Figure 4 is a diagram of a commercial aircraft gas turbine engine simulated by C-MAPSS. The generated dataset has been commonly used in the field of prognostics and health management for performance evaluation (Ramasso and Saxena, 2014).

The C-MAPSS simulates a 90,000 lb engine model with various environmental conditions: throttle resolver angle (TRA, 20-100), flying heights (0-40,000 ft), and Mach number (0-0.90). The C-MAPSS simulation takes 14 variables as input to simulate a set of different scenarios of rotating components. The output from the simulation includes 58 different variables in the form of sensor response surfaces and operability margins, of which 21 variables were used for health condition monitoring

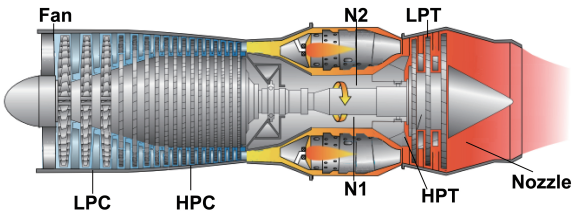

Fig. 4. An illustrative example for aircraft engine.

and prognosis. Table 1 provides a comprehensive description of the C-MAPSS output. The dataset generated for our study consists of 100 historical units for training and 100 in-service units for testing. Each training units have the complete sensor data until the end of their life cycle, while for the testing units, we terminated the sensor-based monitoring at a random time instant before the failure time.

Table 1. Description of the Outputs from the Simulated Dataset

Symbol	Description	Units
T2	Total temperature at fan inlet	°R
T24	Total temperature at LPC outlet	°R
T30	Total temperature at HPC outlet	°R
T50	Total temperature at LPT outlet	°R
P2	Pressure at fan inlet	psia
P15	Total pressure in bypass-duct	psia
P30	Total pressure at HPC outlet	psia
Nf	Physical fan speed	rpm
Nc	Physical core speed	rpm
epr	Engine pressure radio(P50/P2)	–
Ps30	Static pressure at HPC outlet	psia
phi	Ratio of fuel flow to Ps30	pps/psi
NRf	Corrected fan speed	rpm
NRc	Corrected core speed	rpm
BPR	Bypass Ratio	–
farB	Burner fuel-air ratio	–
htBleed	Bleed Enthalpy	–
Nf_dmd	Demanded fan speed	rpm
PCNfR_dmd	Demanded corrected fan speed	rpm
W31	HPT coolant bleed	lbm/s
W32	LPT coolant bleed	lbm/s

Before applying the proposed method, the necessary variable selection is con-

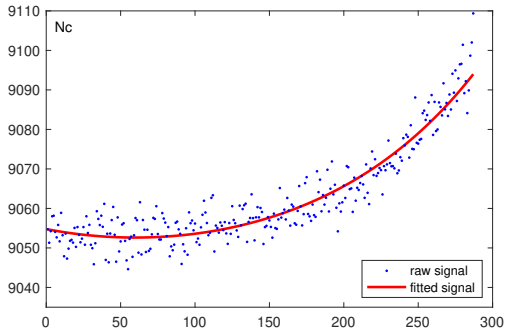

Fig. 5. Representative raw measurements and fitted curves using the exponential function for the five selected signals.

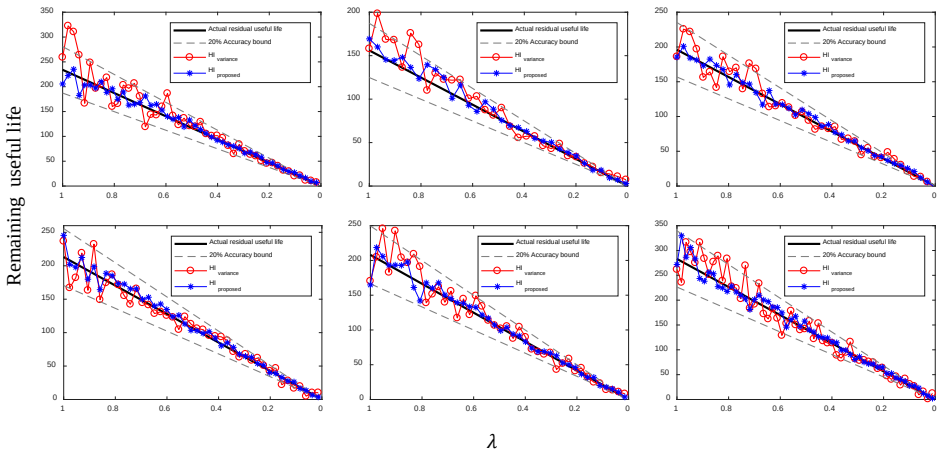

Fig. 6. α-λ performance metric for six randomly selected units.

ducted to avoid information redundancy. In the case study, we adopted the hierarchical clustering algorithm used in (Gao *et al.*, 2020) and identified five most informative clusters for the sensor signals. Since the signals within a cluster are highly correlated, here we arbitrarily selected five sensor signals from each cluster, T50, T24, Nc, htBleed, and T30, for the HI construction. All five selected signals show exponential trends; hence, are suitable for demonstrating the advantageous features of our method. For the illustration purpose, Fig. 5 shows a signal propagation path of an example degradation signal (Nc: physical core speed). Following the literature (Liu *et al.*, 2013), we performed the log-transformation on the sensor measurement data, and then applied the linear mixed effects model with a quadratic

form, i.e., $\boldsymbol{\psi}_t = (1, t, t^2)$, to model the degradation path. Before the neural network training step, the sensor measurements from the C-MAPSS dataset are normalized to be in the range of [0,1] through min-max normalization as:
$x'_{n,s,t} = \frac{x_{n,s,t} - \min(\mathbf{x}_{1,s}, \cdots, \mathbf{x}_{n,s})}{\max(\mathbf{x}_{1,s}, \cdots, \mathbf{x}_{n,s}) - \min(\mathbf{x}_{1,s}, \cdots, \mathbf{x}_{n,s})}$.

4.2. *Model Training and Demonstration of the RUL Prediction*

In the model training process, a proper network structure and penalty coefficients λ_1 and λ_2 need to be determined. As the training dataset contains the full lifetime monitoring data, the cross-validation is performed by calculating the prediction error at all observation time instances, and the mean error is used as the validation error for determining the suitable model structure and for specifying the hyperparameters. The five-fold cross-validation is conducted with 100 training units and the performance is measured by the mean of absolute percentage error (MAPE) defined as:

$$\begin{aligned} \text{MAPE}(\%) &= \frac{100}{N} \sum_{n=1}^{N} \left| \frac{\left(\hat{R}_n + \tau_n\right) - (R_n + \tau_n)}{R_n + \tau_n} \right| \\ &= \frac{100}{N} \sum_{n=1}^{N} \left| \frac{\hat{R}_n - R_n}{T_n} \right|, \end{aligned}$$

where $\widehat{R}_n$ is the predicted RUL, R_n is the true RUL, τ_n is the time of prediction, T_n denotes the entire lifecycle of unit n.

For the RUL prediction, the Bayesian updating method is used to predict the future signal observations of each selected sensor, then the neural network model constructs the composite HI based on the predicted sensor signals. The coefficient k in the loss function is set to be 0.006. The initial learning rate is set to 0.01. The optimization of each neural network structure is repeated 100 times using randomly chosen initial model parameters i.e., weights and biases. Then, the set of model parameters with the smallest fitting loss is selected as the training parameter. The optimal penalty coefficients determined by the cross-validation are $\lambda_1 = 1.7$, and $\lambda_2 = 4.2$.

To study the effect of the proposed minimum variance of the predicted and the observed HIs loss, here we compare the HI developed from the traditional variance loss, i.e., using Constraint 1 as the first term in Equation (8), denoted as $\text{HI}_{\text{variance}}$, with the HI that obtained through the optimization of Equation (8), denoted as $\text{HI}_{\text{proposed}}$. Figure 6 shows the prediction performance of our method using the α-λ metric on six randomly selected units (unit #20, 27, 42, 51, 71, 95). The α-λ metric is a widely-accepted performance metric for evaluating the accuracy of the RUL prediction (Saxena *et al.*, 2008a), where α specifies the error bound on the RUL prediction, i.e., $[(1-\alpha)]R_n \leq \widehat{R}_n \leq [(1+\alpha)]R_n$, and λ represents the percentage of prediction time instant from the actual failure time. Here, we set $\alpha = 20\%$. The RUL prediction results of two HIs are mostly within the 20% error bound and

converge as λ decreases. The RUL prediction results always get better as λ becomes smaller because of the accumulated data as we monitor the unit for a longer period of time (Gao *et al.*, 2020). Besides, compared with $\text{HI}_{\text{variance}}$, $\text{HI}_{\text{proposed}}$ generates more precise RUL prediction, especially in the early and middle periods of degradation. As highlighted earlier, the proposed method considers both the predicted and observed HIs across all time instances, which improves the prognostic performance. In other words, minimizing the variance of the predicted HI in the loss function in addition to minimizing the variance of the observed HI can help with providing accurate RUL prediction, especially for the units in its early stage of life cycle with only a few observations.

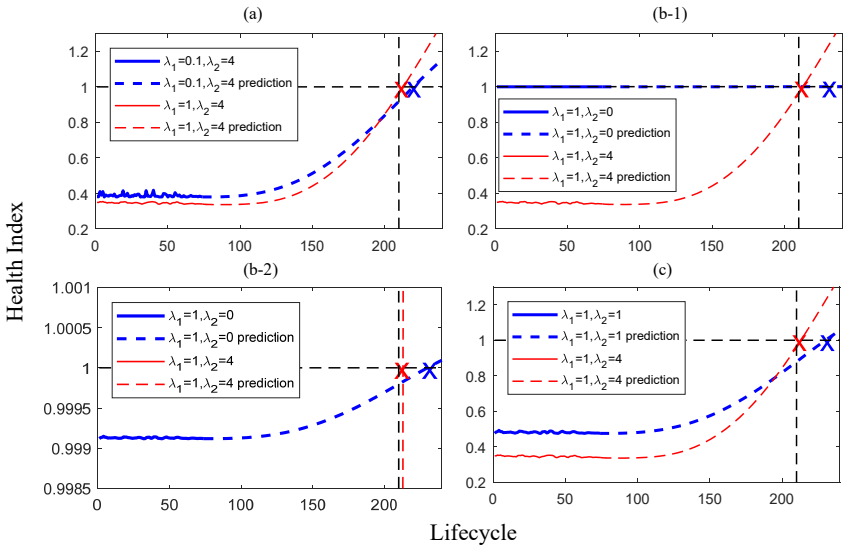

Fig. 7. The influence of increasing the weight of monotonicity and degradation rate constraints in the HI construction.

In the model training step, one of the critical tasks is to find the appropriate penalty parameters. Therefore, investigating the influence of the penalty specifications in the semi-supervised learning process on the performance of the constructed HI is needed. To study this matter, we use a network structure of 5-20-10-5-1 with three different penalty coefficient combinations. A single unit (unit #56) is randomly picked for the analysis and the prediction results are shown in Fig. 7 where the black vertical dashed line indicates the true failure time and the black horizontal dashed line shows the failure threshold. The red and blue solid lines are the observed HI and the dash lines of matching color are the corresponding predicted HI values until the failure time. In Fig. 7, the red lines represent the same scenario where λ_1 and λ_2 are specified as 1 and 4, respectively. The blue lines show the results based

on a set of different values of λ_1 and λ_2. By comparing blue and red lines, we get valuable insights.

Fig. 7(a) shows the impact of reducing the monotonicity weight of the loss function (from $\lambda_1 = 1$ to $\lambda_1 = 0.1$). Reducing λ_1 makes the HI less monotonic and less smooth as shown by the blue solid line. The noisy HI eventually leads to the less accurate RUL prediction. From this result, we see that the HI must have a certain level of monotonicity to ensure satisfactory RUL prediction performance. Our method has Property 2* (the monotonic and accelerated degradation property), whereas the existing HIs in the literature typically only have the monotonic degradation property. The importance of establishing an HI with accelerated degradation property is often overlooked in the literature. However, in some cases, it makes a significant difference in terms of HI quality and prediction accuracy. Fig. 7(b-1) demonstrates the potential issue of having the monotonic degradation property without the accelerated rate property ($\lambda_2 = 0$). There, the blue line is very close to the failure threshold throughout the unit's entire life cycle. To show it more clearly, we rescale the y-axis in Fig. 7(b-2). As we can see, all the observed HI values computed based on the sensor signals (blue solid line) are too high (comparable to the failure threshold), which makes the predicted HI values (blue dashed line) also unreasonable. The variance-based loss function forces the predicted HI at the failure time to approach the failure threshold as closely as possible. On top of that, the monotonicity penalty (λ_1) encourages the HI to be consistent at every prediction time instances. Therefore, the minimizing the variance-based loss while maintaining the monotonicity property can easily lead to an HI similar to the one shown in Fig. 7(b-1) and (b-2), especially for an HI that has a small amount of variation. By introducing the degradation acceleration property in addition to the monotonicity property, we can effectively mitigate the issue. In Fig. 7(c), we make the HI to have an accelerated degradation rate by setting $\lambda_2 = 1$. Although the RUL prediction performance does not improve significantly, the blue lines in Fig. 7 no longer suffer from the issue shown in Fig. 7(b) because the degradation acceleration property increases the signal range.

4.3. *Performance Evaluation and Comparison*

We evaluate the performance of the proposed method and compared it with three existing methods discussed in the Introduction. The benchmark methods are denoted by $\text{HI}_{\text{linear}}$ (Liu *et al.*, 2013), $\text{HI}_{\text{quantile}}$ (Song and Liu, 2018), $\text{HI}_{\text{kernel}}$ (Song *et al.*, 2017) and HI_{llr} (Wang *et al.*, 2024a). Both the $\text{HI}_{\text{linear}}$ and $\text{HI}_{\text{quantile}}$ methods construct the HI based on a linear combination of sensor data, while the $\text{HI}_{\text{kernel}}$ method uses a kernel method to fuse multi-sensor signals accommodating the potential nonlinearity. Therefore, the $\text{HI}_{\text{kernel}}$ method offers more flexibility when it comes to modeling the nonlinearity and is expected to perform better. Unlike the aforementioned parametric models that make specific assumptions about the signal path, HI_{llr} uses a nonparametric representation to ensure flexibility. Besides,

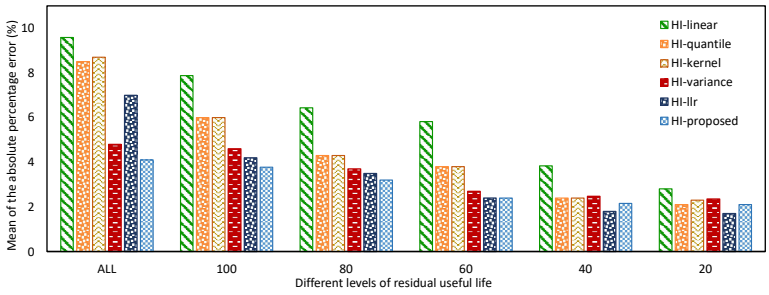

Fig. 8. Comparison of RUL prediction errors using the proposed method and other compared approaches.

this approach directly construct local models to describe the relationship between these features and failure time, which belongs to supervised learning strategy. Our method, denoted by $\mathrm{HI}_{\mathrm{proposed}}$ is also capable of addressing the nonlinearity; hence, it is meaningful to compare the performance of our method to the three existing approaches in the literature and highlight the importance of properly modeling the nonlinear relationship. In addition, we also include another benchmark method based on the conventional variance-based loss function denoted by $\mathrm{HI}_{\mathrm{variance}}$. As previously mentioned, unlike the $\mathrm{HI}_{\mathrm{variance}}$ approach, our method considers both the predicted and the observed HIs.

The results are shown in Fig. 8 where the y-axis represents the MAPE and the x-axis refers the number of testing units with the RUL not exceeding the corresponding days. For example, "ALL" means all the time series for 100 testing units, while other numbers represent testing units with a remaining life not exceeding the corresponding days. From Fig. 8 , we see that the prediction errors of all methods decrease as the unit approaches to its failure time. Compared with three methods $\mathrm{HI}_{\mathrm{linear}}$, $\mathrm{HI}_{\mathrm{quantile}}$ and $\mathrm{HI}_{\mathrm{kernel}}$, the performance of the proposed method is consistently better, especially at the early stage of life cycle. At the later stage (e.g., 40 and 20), all methods show comparable prediction accuracy except for the $\mathrm{HI}_{\mathrm{linear}}$ method primarily due to its simple model structure. The performances of $\mathrm{HI}_{\mathrm{variance}}$ and $\mathrm{HI}_{\mathrm{proposed}}$ are better or comparable to $\mathrm{HI}_{\mathrm{kernel}}$. From this results, we see that using the neural network for nonlinear data fusion is a promising alternative over the kernel-based method. Furthermore, $\mathrm{HI}_{\mathrm{proposed}}$ outperforms $\mathrm{HI}_{\mathrm{variance}}$ method. This result highlights the importance of considering both the predicted and the observed HIs for improving the prediction accuracy. Compared with $\mathrm{HI}_{\mathrm{llr}}$ that focuses on reducing feature uncertainty through importance sampling, our method explicitly considers both the predicted and observed HIs at failure time, leading to more consistent prognostic performance, especially in the early stages of degradation.

5. Conclusion and Discussion

To correctly assess the health condition of various engineering systems, diverse condition monitoring signals are collected by multiple sensors. However, most health assessment approaches either fail to incorporate the actual failure information in the HI construction process or show limitations on characterizing the complex nonlinear nature of multi-sensor signals. To address this issue, we propose a semi-supervised nonlinear data fusion approach for health assessment with the prognostic accuracy of the RUL considered. In this method, a neural network structure is established to model the highly nonlinear relationship between the sensor signals and the underlying health condition. The variances of both the predicted and observed HIs at the failure time are considered in loss function to improve the prognostic accuracy. By carefully designing the loss function, the monotonic accelerated degradation property is obtained. A customized Adam algorithm specifically tailored to our method is developed to efficiently estimate the model parameters. Last, for the RUL prediction, a mixed effects model with the Bayesian updating framework is adopted to accurately predict the future signal for each in-service unit. The proposed health assessment framework is evaluated using the C-MAPSS dataset. From the performance evaluation results, the satisfactory RUL prediction performance of the proposed method is reported and, from the performance comparison with the existing approaches in the literature, we show that the proposed method has a great potential. Altogether, we believe this study provides a promising alternative based on the semi-supervised nonlinear learning scheme for the HI construction.

Despite the meaningful results reported in this paper, there are some limitations. First, our method considers only a single failure mode and a single operating condition. Extending the model to a more general case possibly with multiple failure modes and varying degrees of operating condition will be highly appreciated especially by practitioners. Second, the established model for a unit can be extended to more complex mechanical system by considering the physical interation between each component through graphic model. We shall investigate along these lines and report the results in the future.

Statements and Declarations

No potential conflict of interest was reported by the authors.

Funding

This study was supported by the National Natural Science Foundation of China under Grant NSFC-71932006 and Grant NSFC-72171003.

Appendix A. Proofs of Statements

We herein present all proofs of the equations presented in this paper.

A.1. *Derivation of Equation (3.4)*

Suppose that $\pi\left(\boldsymbol{\beta}, \sigma^2\right) = \mathrm{IG}\left(\sigma^2 | a_1, a_2\right) N\left(\boldsymbol{\mu}_0, \sigma^2 \boldsymbol{\Sigma}_0\right)$, $\boldsymbol{\beta}$ is of dimension q, then we can derive

$$
\begin{aligned}
& p\left(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{x}_{1:t}\right) \propto p\left(\boldsymbol{\beta}, \sigma^2\right) p\left(y_{1:t} | \boldsymbol{\beta}, \sigma^2\right) \\
& \propto\left[\frac{1}{2 \pi\left|\sigma^2 \boldsymbol{\Sigma}_0\right|^{\frac{1}{2}}} e^{-\frac{\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)}{2 \sigma^2}}\right] \\
& \times\left[\frac{a_2^{a_1}}{\Gamma\left(a_1\right)}\left(\sigma^2\right)^{-a_1-1} e^{-\frac{a_2}{\sigma^2}}\right] \\
& \times\left[(2 \pi)^{-\frac{t}{2}}\left(\sigma^2\right) e^{\frac{-\left|y_{1:t}-\boldsymbol{\Psi}_{1, t} \boldsymbol{\beta}\right|^2}{2 \sigma^2}}\right] \\
& \propto \frac{\left(\sigma^2\right)^{-a_1-1-\frac{t}{2}}}{2 \pi\left(\sigma^2\right)^{\frac{q}{2}}} \times \exp\left[-\frac{\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)}{2 \sigma^2}\right] \\
& \times \exp\left[-\frac{2 a_2+\left|y_{1:t}-\boldsymbol{\Psi}_{1: t} \boldsymbol{\beta}\right|^2}{2 \sigma^2}\right] \\
& \propto\left(\sigma^2\right)^{-a_1-\frac{t}{2}-1} \times \exp\left[-\frac{y_{1: t}^{\mathrm{T}} y_{1: t}+2 a_2+\boldsymbol{\mu}_0^T \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0}{2 \sigma^2}\right] \\
& \times \exp\left[\frac{\boldsymbol{\mu}_t^{\mathrm{T}}\left(\boldsymbol{\Psi}_{1: t}^{\mathrm{T}} \boldsymbol{\Psi}_{1: t}+\boldsymbol{\Sigma}_0^{-1}\right) \boldsymbol{\mu}_t}{2 \sigma^2}\right] \\
& \times \frac{1}{\left(\sigma^2\right)^{q / 2}} \exp\left[-\frac{\left(\boldsymbol{\beta}-\boldsymbol{\mu}\right)^{\mathrm{T}}\left(\boldsymbol{\Psi}_{1: t}^{\mathrm{T}} \boldsymbol{\Psi}_{1: t}+\boldsymbol{\Sigma}_0^{-1}\right)\left(\boldsymbol{\beta}-\boldsymbol{\mu}_t\right)}{2 \sigma^2}\right] \\
& \propto \mathrm{IG}\left(a_1+\frac{t}{2}, a_2+\frac{H_{1, t}}{2}\right) \times N\left(\boldsymbol{\mu}_t, \sigma^2 \boldsymbol{\Sigma}_t\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\Sigma}_t &= \left(\boldsymbol{\Psi}_{1, t}^{\mathrm{T}} \boldsymbol{\Psi}_{1, t}+\boldsymbol{\Sigma}_0^{-1}\right)^{-1}, \\
\mathbf{M}_t &= \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0+\boldsymbol{\Psi}_{1, t} y_{1: t}, \\
\boldsymbol{\mu}_t &= \boldsymbol{\Sigma}_t \mathbf{M}_t, \\
H_t &= \left|\boldsymbol{x}_{1: t}\right|^2+\boldsymbol{\mu}_0^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0-\mathbf{M}_t^{\mathrm{T}} \boldsymbol{\Sigma}_t \mathbf{M}_t.
\end{aligned}
$$

A.2. *Derivation of Equation (3.4)*

Based on Equation 3.4, since

$$
\hat{\boldsymbol{x}}_{t+1: t+L}=\boldsymbol{\Psi}_{t+1, t+L} \boldsymbol{\beta}+\sigma \varepsilon_{t+1: t+L},
$$

then we can have

$$
\left(\hat{\boldsymbol{x}}_{t+1: t+L} | \sigma^2, \boldsymbol{x}_{1: t}\right) \sim N\left(\boldsymbol{\Psi}_{t+1, t+L} \boldsymbol{\mu}_t, \sigma^2\left(\boldsymbol{I}+\boldsymbol{\Psi}_{t+1, t+L} \boldsymbol{\Sigma}_t \boldsymbol{\Psi}_{t+1, t+L}^{\mathrm{T}}\right)\right).
$$

Let $\boldsymbol{\mu}_* = \boldsymbol{\Psi}_{t+1,t+L}\boldsymbol{\mu}_t$, $\boldsymbol{\Sigma}_* = \boldsymbol{I} + \boldsymbol{\Psi}_{t+1,t+L}\boldsymbol{\Sigma}_t\boldsymbol{\Psi}_{t+1,t+L}^{\mathrm{T}}$,

$$
\begin{aligned}
p\left(\hat{\boldsymbol{x}}_{t+1:t+L}|\boldsymbol{x}_{1:t}\right) &= \int p\left(\hat{\boldsymbol{x}}_{t+1:t+L}|\sigma^2,\boldsymbol{x}_{1:t}\right) p\left(\sigma^2|\boldsymbol{x}_{1:t}\right) d\sigma^2 \\
&\propto \int \left(\sigma^2\right)^{-\frac{L}{2}} \left|\boldsymbol{\Sigma}_*\right|^{-\frac{1}{2}} \times \exp\left[-\frac{\left(\hat{\boldsymbol{x}}_{t+1:t+L}-\boldsymbol{\mu}_*\right)\boldsymbol{\Sigma}_*^{-1}\left(\hat{\boldsymbol{x}}_{t+1:t+L}-\boldsymbol{\mu}_*\right)^{\mathrm{T}}}{2\sigma^2}\right] \\
&\times \left(\sigma^2\right)^{-a_1-\frac{t}{2}-1} \exp\left[-\frac{2a_2+H_t}{2\sigma^2}\right] d\sigma^2 \\
&\propto \int \left(\sigma^2\right)^{-a_1-\frac{t+L}{2}-1} \times \exp\left[-\frac{2a_2+H_t}{2\sigma^2}\right] \\
&\times \exp\left(-\frac{\left(\hat{\boldsymbol{x}}_{t+1:t+L}-\boldsymbol{\mu}_*\right)\boldsymbol{\Sigma}_*^{-1}\left(\hat{\boldsymbol{x}}_{t+1:t+L}-\boldsymbol{\mu}_*\right)^{\mathrm{T}}}{2\sigma^2}\right) d\sigma^2 \\
&\propto \frac{\Gamma\left(a_1+\frac{t+L}{2}\right)}{\left[\frac{\left(\hat{\boldsymbol{x}}_{t+1:t+L}-\boldsymbol{\mu}_*\right)\boldsymbol{\Sigma}_*^{-1}\left(\hat{\boldsymbol{x}}_{t+1:t+L}-\boldsymbol{\mu}_*\right)^{\mathrm{T}}+(2a_2+H_t)}{2}\right]^{a_1+\frac{t+L}{2}}} \\
&\propto \left[1+\frac{1}{v}\frac{\left(\hat{\boldsymbol{x}}_{t+1:t+L}-\boldsymbol{\mu}_*\right)\boldsymbol{\Sigma}_*^{-1}\left(\hat{\boldsymbol{x}}_{t+1:t+L}-\boldsymbol{\mu}_*\right)^{\mathrm{T}}v}{2a_2+H_t}\right]^{-\frac{L+v}{2}}
\end{aligned}
$$

where $v = 2a_1 + t$. Therefore, the future observation vector $\hat{\boldsymbol{x}}_{t+1:t+L}$ given the historical observations up to time t follows a multivariate t-distribution:

$$
\hat{\boldsymbol{x}}_{t+1:t+L}|\boldsymbol{x}_{1:t} \sim MT\left(2a_1+t, \boldsymbol{\Psi}_{t+1,t+L}\boldsymbol{\mu}_t, \frac{2a_2+H_t}{2a_1+t}(\boldsymbol{I}+\boldsymbol{\Psi}_{t+1,t+L}\boldsymbol{\Sigma}_t\boldsymbol{\Psi}_{t+1,t+L}^{\mathrm{T}})\right).
$$

References

Bontempi, G, SB Taieb and Y Borgne (2013). *Machine learning strategies for time series forecasting*, Business Intelligence: Berlin, Heidelberg.

Bottou, L (2010). Large-scale machine learning with stochastic gradient descent, in *Proceedings of COMPSTAT'2010*, Springer, pp. 177–186.

Cai, B, H Fan, X Shao, Y Liu, G Liu, Z Liu and R Ji (2021). Remaining useful life re-prediction methodology based on wiener process: Subsea christmas tree system as a case study, *Computers & Industrial Engineering* 151, 106983.

Chen, L, G Xu, S Zhang, W Yan and Q Wu (2020). Health indicator construction of machinery based on end-to-end trainable convolution recurrent neural networks, *Journal of Manufacturing Systems* 54, 1–11.

Chen, T and H Chen (1995). Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems, *IEEE Transactions on Neural Networks* 6(4), 911–917.

Franklin, J (2005). The elements of statistical learning: data mining, inference and prediction, *The Mathematical Intelligencer* 27(2), 83–85.

Gao, Y, Y Wen and J Wu (2020). A neural network-based joint prognostic model for data fusion and remaining useful life prediction, *IEEE transactions on neural networks and learning systems* 32(1), 117–127.

Guo, L, N Li, F Jia, Y Lei and J Lin (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings, *Neurocomputing* 240, 98–109.

Hong, S, Z Zhou, C Lu, B Wang and T Zhao (2015). Bearing remaining life prediction using gaussian process regression with composite kernel functions, *Journal of Vibroengineering* 17(2), 695–704.

Jain, M and K Priya (2005). Software reliability issues under operational and testing constraints, *Asia-Pacific Journal of Operational Research* 22(01), 33–49.

Kingma, DP and J Ba (2014). Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* .

Liu, K, NZ Gebraeel and J Shi (2013). A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis, *IEEE Transactions on Automation Science & Engineering* 10(3), 652–664.

Liu, R, B Yang and AG Hauptmann (2019a). Simultaneous bearing fault recognition and remaining useful life prediction using joint-loss convolutional neural network, *IEEE Transactions on Industrial Informatics* 16(1), 87–96.

Liu, Y, X Hu and W Zhang (2019b). Remaining useful life prediction based on health index similarity, *Reliability Engineering System Safety* 185, 502–510.

Miao, H, B Li, C Sun and J Liu (2019). Joint learning of degradation assessment and rul prediction for aeroengines via dual-task deep lstm networks, *IEEE Transactions on Industrial Informatics* 15(9), 5023–5032.

Nelson, W (1990). Accelerated testing: Statistical models, test plans, and data analysis, *Technometrics* 33(2), 236–238.

Ramasso, E and A Saxena (2014). Review and analysis of algorithmic approaches developed for prognostics on cmapss dataset, in *Annual Conference of the Prognostics and Health Management Society 2014*.

Saxena, A, J Celaya, E Balaban, K Goebel, B Saha, S Saha and M Schwabacher (2008a). Metrics for evaluating performance of prognostic techniques, in *2008 international conference on prognostics and health management*, IEEE, pp. 1–17.

Saxena, A, K Goebel, D Simon and N Eklund (2008b). Damage propagation modeling for aircraft engine run-to-failure simulation, in *2008 international conference on prognostics and health management*, IEEE, pp. 1–9.

Sibi, P, SA Jones and P Siddarth (2013). Analysis of different activation functions using back propagation neural networks, *Journal of theoretical and applied information technology* 47(3), 1264–1268.

Song, C and K Liu (2018). Statistical degradation modeling and prognostics of multiple sensor signals via data fusion: A composite health index approach, *IISE Transactions* 50(10), 853–867.

Song, C, K Liu and X Zhang (2017). Integration of data-level fusion model and kernel methods for degradation modeling and prognostic analysis, *IEEE Transactions on Reliability* 67(2), 640–650.

Taşcı, B, A Omar and S Ayvaz (2023). Remaining useful lifetime prediction for predictive maintenance in manufacturing, *Computers & Industrial Engineering* 184, 109566.

Wang, D, A Wang and C Song (2024a). Flexible degradation modeling via the integration of local models and importance sampling, *IEEE Transactions on Instrumentation and Measurement* .

Wang, H, X Li, Z Zhang, X Deng and W Jiang (2024b). A deep learning based health index construction method with contrastive learning, *Reliability Engineering System Safety* 242, 109799.

Wei, Y, D Wu and J Terpenny (2021). Learning the health index of complex systems using dynamic conditional variational autoencoders, *Reliability Engineering System*

Safety 216, 108004.

Wen, Y, J Wu, D Das and T Tseng (2018a). Degradation modeling and rul prediction using wiener process subject to multiple change points and unit heterogeneity, *Reliability Engineering & System Safety* 176, 113–124.

Wen, Y, J Wu, Q Zhou and T Tseng (2019). Multiple-change-point modeling and exact bayesian inference of degradation signal for prognostic improvement, *IEEE Transactions on Automation Science and Engineering* 16(2), 613–628.

Wen, Y, J Wu, Q Zhou and TL Tseng (2018b). Multiple-change-point modeling and exact bayesian inference of degradation signal for prognostic improvement, *IEEE Transactions on Automation Science and Engineering* 16(2), 613–628.

Yu, JB and S Wang (2009). Using minimum quantization error chart for the monitoring of process states in multivariate manufacturing processes, *Computers & Industrial Engineering* 57(4), 1300–1312.

Yuan, Y, Q Yang, G Wang, J Ren, Z Wang, F Qiu, K Li and H Liu (2025). Combined improved tuna swarm optimization with graph convolutional neural network for remaining useful life of engine, *Quality and Reliability Engineering International* 41(1), 174–191.

Zhang, J, N Jiang, H Li and N Li (2019). Online health assessment of wind turbine based on operational condition recognition, *Transactions of the Institute of Measurement and Control* 41(10), 2970–2981.

Zhang, J, P Wang, R Yan and RX Gao (2018). Long short-term memory for machine remaining life prediction, *Journal of manufacturing systems* 48, 78–86.

Zhou, Q, J Son, S Zhou, X Mao and M Salman (2014). Remaining useful life prediction of individual units subject to hard failure, *IIE Transactions* 46(10), 1017–1030.