Remaining Useful Life Interval Prediction for Complex System Based on BiGRU Optimized by Log-Norm

XIAOJIA YAN, WEIGE LIANG, AND DONGXUE XU
College of Weapon Engineering, Naval University of Engineering, Wuhan 430033, China
Corresponding author: Weige Liang (lwg_nue@outlook.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61640308, and in part by the Natural Science Foundation of Hubei Province under Grant 2019CFB362.

ABSTRACT The task of remaining useful life (RUL) uncertainty management is the major challenge in solving the failure of the complex mechanical system. Primary research methods use statistical models or stochastic processes to fit the distribution of historical degradation data. However, it is difficult to accurately capture the degradation information of monitoring big data through statistics in practice. In this paper, the prediction interval (PI) obtained by the proposed feature attention-log-norm bidirectional gated recurrent unit (FA-LBiGRU) model is adopted to quantify the prediction uncertainty of RUL. Initially, the critical feature vectors are extracted from multi-dimensional, nonlinear, and large-scale sensor signals using the feature attention mechanism. Additionally, the BiGRU network is used to model and learn the time-varying characteristics of the attention-weighted features from the forward and backward directions, and the network parameters are trained by the maximum log-likelihood loss function. Ultimately, the probability density function based on the lognormal distribution is calculated to measure the uncertainty of the equipment RUL. The effectiveness of the proposed method is verified through the well-known benchmark data set of the turbofan engines provided by NASA. The experimental results show that the proposed methods can obtain higher point prediction accuracy for the complex system compared with state-of-the-art approaches and high-quality PIs satisfying real-time requirements.

INDEX TERMS Fusion model, gated recurrent unit, prediction intervals, remaining useful life, system prognostics, uncertainty management.

I. INTRODUCTION
The nonlinear coupling between systems and the uncertainty in the operation process increases the possibility of complex mechanical systems degradation leading to failure, as the structure and function of complex mechanical equipment grow into increasingly abundant ones. As one of the most challenging technologies to address these difficulties, the prediction of Remaining Useful Life (RUL) is an essential basis for evaluating the reliability of mechanical systems [1], ensuring the safe and reliable operation of significant equipment [2], reducing the massive cost of scheduled maintenance [3] and provide technical support for predictive maintenance [4].

RUL prediction forecasts the remaining time for equipment to decline to the failure threshold by analyzing mechanical equipment’s historical performance degradation process and dynamically anticipating future changes in equipment health state. Three prediction methods for prognostics are mainly: model-based, data-driven, and hybrid methods [5]. The model-based and hybrid methods can meticulously depict the system degradation process and establish a precise mathematical model derived from known mechanics principles to examine the equipment performance degradation process [6]. However, the sizeable mechanical system represented by the turbine engine [7] has a complex mechanical
structure, poor task environment, and diverse failure modes, leading to an extremely high cost to establish an accurate physical failure model and, thus, a challenge to deploy in practical application. Compared with the above methods, the data-driven methods can sufficiently dig up historical monitoring sensor data during machinery operation, constructing a RUL prediction model from derived degradation features. Under such circumstances, data-driven methods have become a research hotspot with the advantage that they can refrain from prior knowledge requirements and the modelling process of the complex mechanism [8], [9], [10].

Data-driven prediction models are mainly divided into Machine Learning (ML) and statistical models. The ML methods represented by Deep Learning (DL) realize end-to-end RUL prediction by establishing the mapping relationship between monitoring data and failure time. Recurrent Neural Network (RNN) under DL, particularly Long-Short Term Memory (LSTM), presented by Mamo and Wang [11], and Gated Recurrent Unit (GRU), presented by He et al. [12], is suitable for processing time series via remembering historical information in the monitoring data. Also, they can solve the problems such as unobvious long-term features due to traditional shallow learning and gradient disappearance. Zhang et al. [13] investigated lithium-ion battery RUL prediction based on LSTM to solve the problem that existing techniques cannot effectively learn the long-term dependencies among capacity degradations. Reference [14] introduced a transfer learning method based on bidirectional GRU (BiGRU) to adapt RUL estimation of bearings under different working conditions, which efficiently recognized different running states of bearings and obtained corresponding training labels. In [15], LSTM is utilized for prognostics with the degradation sequence of equipment, which takes the operation state indicator and concatenated equipment characteristics as input. In short, the DL methods can thoroughly grasp the information contained by time series to more accurately predict RUL without the exact physical model and pre-knowledge. Nevertheless, these methods can only provide point-wise estimation rather than interval estimation of RUL, which is challenging to quantify the probability distribution of uncertainty in the RUL prediction process and increase the risk of subsequent maintenance decisions in health management.

In contrast, the statistical method utilizes a statistical model or random process to distribute the fitting of historical equipment failure data. The probability density function (PDF) of RUL is constructed using parameter estimation, and then the prediction intervals (PIs) of RUL are achieved by carrying out statistical analysis of real-time sensor data. Typically, to take nonlinearity and uncertainty of RUL prediction into account, representative univariate statistical methods, e.g., wiener and gamma degradation process, are developed via parameters updating of probability distribution [16], [17], [18]. Nonetheless, with the rapid development of sensors and information technology, various networking sensors are usually used to grasp multi-source data to describe the operating status of machinery comprehensively. Performance degradation indicators that characterize system health status are often not unique, showing the features of multiple degradation variables. Hence, to further constructs a 1-D health index (HI), the method of transforming multidimensional data into one-dimensional data has been widely studied. Lyu et al. [19] and Li et al. [20] proposed data fusion methods via automatically selecting and combining multiple sensor signals to better characterize the degradation process. In [21], to address the drawbacks of sixteen original features without better trendability, Principal Component Analysis (PCA) is used for dimensionality reduction. Although information fusion methods can be implemented in the traditional univariate degradation modelling, it is tough to describe the coupling relationship of multidimensional data. Also, the degradation information of monitoring data is prone to be utilized insufficiently.

Therefore, the fusion RUL prediction methods based on ML and statistical models have been widely discussed by global scholars. For instance, the combination of a deep belief network and particle filter can forecast the uncertainty of RUL [22]. In [23], an improved echo state Gaussian process, which is able to provide an estimation of prediction uncertainty, is developed as a novel data-driven approach for RUL prediction and PIs construction. Moreover, Peng et al. proposed a Bayesian deep-learning-based (BDL-based) method for health prognostics with uncertainty quantification in [24]. Hu et al. [25] proposed a RUL prediction model based on the deep belief network for feature extraction and diffusion process for constructing PIs. Liao et al. [26] obtained the prediction uncertainty based on a classical bootstrapping approach with LSTM. Although the above methods can combine the advantages of both ML and statistical methods, it still exists following challenges that impede the accuracy and efficiency of RUL interval prognosis in mechanical systems:

1) The above studies in the literature can only determine the optimal index of multidimensional degradation characteristics through the index filtration technology for the random degradation model, i.e. monotonicity, trendability, and prognosability [27], in which the risk of missing critical information in conversion from nonlinear to linear is exposed.

2) The deep network feature extraction is independent of the stochastic model of constructing the degradation process, thus former is hard to match the latter.

Consequently, this article, aiming to concomitantly settle these two challenges, proposes a novel end-to-end approach, named feature attention-log-norm bidirectional gated recurrent unit (FA-LBiGRU) model, which is presented to quantify the predictive uncertainty of mechanical systems. The main contributions of this paper are summarized as follows:

1) The feature attention (FA) mechanism is utilized to withdraw key feature vectors from multidimensional, nonlinear, and large-scale sensor signals in order to realize the dynamic change of input features.
2) A log-norm bidirectional gated recurrent unit (LBiGRU) is developed for a fusion of ML and statistical models to construct PIs while predicting RUL. More concretely, using the BiGRU network to study the time-varying characteristics of attention-weighted features from forwarding and backward directions. Moreover, due to its flexibility, a novel maximum log-likelihood loss function based on the lognormal distribution (LN) is framed to train the network parameters.

The remainder of the paper is organized as follows. Section II presents the theoretical background of the proposed methodology. In Section III, the proposed framework and its operation process are revealed. A case study on the C-MAPSS dataset is performed in Section IV, which is utilized to demonstrate the validity and predominance of the proposed approach. Finally, the work is summarized and concluding remarks are drawn in Section V.

II. THEORETICAL BACKGROUND
A. FEATURE ATTENTION MECHANISM
In order to more comprehensively evaluate health status and performance degradation trend, monitoring data with multiple features often require comprehensive consideration for mechanical system working in complex environments. Further, these features contain diverse contributions to RUL predictions. Hence, FA mechanism, whose specific structure is shown in Fig. 1, is applied to select the key features of the original input data, which can effectively reflect the correlation between features and RUL prediction, and improve the feature extraction ability of the prediction model.

![FIGURE 1. Illustration of feature attention.](image)

In Fig. 1, $T$ represents the time length of the time series, and $F$ represents the features length. Firstly, the mean value of each time sample segmented by sliding window is obtained through Global Average Pooling (GAP) of the feature dimension, which is compressed from $T \times F$ to $1 \times F$, and the calculation process can be expressed as

$$y^f = \frac{1}{T} \sum_{t=1}^{T} x^f_t \quad (1 \leq f \leq F, 1 \leq t \leq T) \quad (1)$$

where, $y^f$ is the vector obtained through GAP, $g(\cdot)$ represents the global average pooling operation of feature dimension, $X^f$ is data at all time under $f^{th}$ feature, $x^f_t$ is the data at $t^{th}$ time of the $f^{th}$ feature.

Then, the attention weights corresponding to each feature are quantified through the two fully connected layers, and the weights are normalized by sigmoid activation function. Its calculation function is

$$a^f = \sigma(W_2 \delta(W_1 y^f + b_1) + b_2) \quad (2)$$

where, $W_1, b_1$ and $W_2, b_2$ are the weights and offsets of two layers respectively, $\delta$ is the ReLU activation function, $\sigma$ is the sigmoid activation function.

Finally, normalized attention weight $a^f$ multiplied by the corresponding input data $x$ to enhance or weaken the $f^{th}$ eigenvector, and a new attention weighted input $x'$ is obtained as follows:

$$x' = a^f \odot x = [a_1 x^1, a_2 x^2, \cdots, a_F x^F] \quad (3)$$

B. BIDIRECTIONAL GATED RECURRENT UNIT

The main way to predict the RUL is to learn the spatio-temporal information of performance degradation from the sensor data of mechanical system. LSTM can make full use of time correlation of data to process time series data. However, its complex internal structure leads to long training time, which is not conducive to practical application deployment [28].

GRU optimizes LSTM by combining forgetting gates and input gates to reduce the number of gates and improve the training efficiency of the network while ensuring memory ability. Memory unit is the core of GRU network, which is composed of update gate, reset gate and candidate hidden layer. Its network topology is shown in Fig. 2.

![FIGURE 2. Structure of GRU network.](image)

It can be seen from Fig. 2 that $h_{t-1}$, $h_t$ and $h_{t+1}$ are the hidden states at time $t-1$, $t$ and $t+1$ respectively, representing the memory information of the input sequence. $x_{t-1}$, $x_t$ and $x_{t+1}$ are the input of memory units at moments $t-1$, $t$ and $t+1$ respectively. In addition, the arrow indicates the direction of information flow. The specific calculation process of each GRU memory unit is

$$\begin{align*}
    r_t &= \sigma(W_r [x_t, h_{t-1}] + b_r) \\
    z_t &= \sigma(W_z [x_t, h_{t-1}] + b_z) \\
    n_t &= \tanh(W_n [x_t, r_t \times h_{t-1}] + b_n) \\
    h_t &= (1 - z_t) \times h_{t-1} + z_t \times n_t
\end{align*} \quad (4)$$
where, $W_r$ and $b_r$, $W_z$ and $b_z$, $W_f$ and $b_f$ are the weight matrix and bias vector of reset gate $r_i$, update gate $z_i$ and candidate hidden layer $n_i$, respectively. Furthermore, $r_i$ is responsible for controlling the memory that needs to be retained in the transmission process of the unit. When $r_i$ approaches 1, $n_i$ mainly contains the information composed of the current input $x_i$ and the hidden state $h_{i-1}$ of the previous unit input. When $r_i$ approaches 0, then $n_i$ contains mainly $x_i$. Inversely, $z_i$ controls the degree of selective abandonment of $h_{i-1}$. When $z_i$ approaches 1, the memory unit mainly outputs $h_{i-1}$ of the previous moment. When $z_i$ approaches 0, the memory unit mainly outputs $n_i$, that is, $h_i$ at the current moment.

In summary, GRU gets two key gating states through the hidden state $h_{i-1}$ of the previous unit and the input $x_i$ of the current moment. The output vectors of $r_i$ and $z_i$ determine which information can be used as the final output of GRU. In order to deepen the level of feature extraction of the original time series and further improve the accuracy of the model output, two independent GRUs with different directions are superimposed together to form a BiGRU network. The specific structure is shown in Fig. 3.

![FIGURE 3. Structure of BiGRU network.](image)

The calculation process is:

$\mathbf{h} = f(\mathbf{h}_{f}, \mathbf{h}_{b})$ (5)

where, $f(\cdot)$ is the mapping function of the full connection layer.

**C. LOGNORMAL DISTRIBUTION**

The reliability of mechanical equipment is usually measured by probability index, and its life is a non-negative variable, which is difficult to fit its trend through normal distribution. However, LN is one of the commonly used life distributions in the field of reliability, which can convert the life data with large difference in values into standard normal distribution after taking logarithm. Suppose that a random variable $Y = \ln(X)$ obeys LN with mean $\mu$ and variance $\sigma$, then the random variable $X$ satisfies the LN distribution with parameter ($\mu$, $\sigma^2$). The formulas for calculating the statistical characteristics are shown in Table 1.

**TABLE 1. Statistical characteristics of lognormal distribution.**

| Statistical characterization | Computing formula |
|----------------------------|-------------------|
| Probability density function| $\frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln(x) - \mu)^2}{2\sigma^2}}$ |
| Mean value                  | $e^{\frac{\mu+\sigma^2}{2}}$ |
| Median value                | $e^\mu$             |
| Mode value                  | $e^{(\mu-\sigma^2)}$ |

The distribution of statistical eigenvalues of LN is shown in Fig. 4-a, where the PDF of mode value is the largest. Furthermore, LN shows negatively skewness owing to the sample is mainly concentrated on the left side of the mean value. Therefore, LN is suitable for describing the probability distribution of RUL in the degradation process, since the predicted RUL should be less than the real RUL to ensure the safety requirements of the equipment. Fig. 4-b and Fig. 4-c show the influence of different $\mu$ and $\sigma$ on PDF. The smaller $\mu$ is, the more positive the skewness of LN is, while the smaller $\sigma$ is, the more LN tends to be normal distributed. Due to equipment safety requirements, not timely maintenance will cause disastrous consequences, so early prediction is preferable to delayed prediction. However, premature maintenance is sometimes costly. So, the trend of model prediction requires the choice of $\mu$ and $\sigma$ in a reasonable range. The parameter $\mu$ of PI should be close to the actual value, so the prediction accuracy is higher. While the smaller $\sigma$ represents a narrower PI, so the narrow prediction interval can better measure uncertainty. To sum up, LN can be changed into different shapes only according to two parameters, which is a more flexible distribution and is conducive to describing the life samples with roughly symmetric or right skewness.

**III. PROPOSED METHODOLOGY**

The proposed method consists of offline modelling and online application phase, which are described in Fig. 5. First, the offline modelling phase, in which a FA-LBiGRU model is trained on run-to-failure data composed of measurements of healthy and degraded states of the single mechanical system, is applied to establish the mapping between historical condition monitoring (CM) data and RUL distribution. On account of multidimensional historical time series are the main carrier of degradation information, $N$ systems in the same operating environment are assumed to be monitored by $F$ sensors for each system. Then, CM data whose lifetime is divided into $T_i$ by sliding window is measured for the system $i$, $i = 1, 2, \cdots, N$, which can be represented by matrix $X_i = \{x_1^i, x_2^i, \cdots, x_f^i\}$, $X_i \in \mathbb{R}^{F \times T_i}$, where $x^i = [x_1^i, x_2^i, \cdots, x_f^i]$ is a vector in $X_i$ measured.
at time $t$ of sensor $f$, $t \in [1, T_i], f \in [1, F]$. In addition, RUL labels corresponding to $X_i$ can be expressed as $Y_i = [y^1, y^2, \cdots, y^{T_i}]$, before integration of run-to-failure data to historical CM data set $D = \{X_i, Y_i\}$. During the training period, FA-LBiGRU model takes $D$ to find the optimal parameter $\theta_i$ to maximize likelihood function of RUL, where $\theta_i = [\theta^1_i, \theta^2_i, \cdots, \theta^1_f, \cdots, \theta^F_i]$ depicts RUL distribution most accurately. If the RUL distribution follows LN, $\theta^0_i$ contains two parameters of $(\mu^0_i, \sigma^0_i)$, where $\mu^0_i$ is the mean value and $\sigma^0_i$ is the standard deviation derived from modelling stage of the system $i$ at time $t$.

The construction result of system’s PIs is obtained in the second phase, the online application phase, after handing field CM data into the trained FA-LBiGRU model, which translates the sequence of sensor signals into a couple of new practical predictive values. For effective adaptation of proposed model’s data structure, the CM data and the homologous labels need a preprocessing step corresponding to the first half of the first phase, whose details are described in Section A. The framework of the proposed FA-LBiGRU model for construction of mechanical system’s PIs is then demonstrated in Section B.
A. DATA PREPROCESSING

1) DATA SELECTION (DS)

The construction of feature sets are extracted of various types of CM data, e.g., vibration, pressure, temperature and flow, obtained by diverse sensors installed on mechanical system. However, not all of these data are sensitive to the degradation process, or can provide sufficient profitable information for the establishment of PIs. Hence, to further identify a data subset that recognizes essential roles from independent ones and thus raises the precision of prediction results, three standard feature selection metrics in [27] are inherited and exploited to initially pick useful and representative sensors in this paper. In practical engineering cases, the nonlinear correlation among features is common, yet not considered via these selection metrics. Therefore, FA mechanism in Section B is proposed to overcome this drawback.

2) DATA NORMALIZATION (DN)

Since the multi-dimensional monitoring data have different dimensions, normalization preprocessing must be carried out before the model construction to eliminate the influence of feature consistency. Among various popular normalization techniques, this paper chooses the minimum-maximum normalization method to unify the data to the range of [0, 1]. For the given data set \( x^j_i \), where \( i \) is \( j^{th} \) measurement values of the \( j^{th} \) sensor, \( x^j_{\text{max}} \) and \( x^j_{\text{min}} \) are the maximum and minimum of the measurement from the \( j^{th} \) sensor, then the normalized data set \( \hat{x}^j_i \) is presented as

\[
\hat{x}^j_i = \frac{x^j_i - x^j_{\text{min}}}{x^j_{\text{max}} - x^j_{\text{min}}} \tag{6}
\]

3) RUL LABEL RECTIFICATION

Due to the system being in good condition at the beginning of the operation, the degradation degree is small. At the same time, its performance will decline sharply with the extension of time at the end of service. Suppose the label of monitoring data before the rapid deterioration of engine is set to the total operating cycle minus the number of current operational cycles. In that case, the lag of RUL prediction results will increase. Therefore, it can be considered that the RUL remains unchanged before the engine starts to degrade rapidly, that is, to set a threshold for the RUL label of the training set to make it a piecewise linear function. The research [29] shows that it is better to rectify the critical value of RUL label mutation in the training set to the 130th operation cycle, and the RUL label setting results are shown in Fig. 6.

4) SLIDING WINDOW PRETREATMENT (SWP)

In order to use the limited time series to dig deep into the degradation law, and transform the time series into a three-dimensional input format that the BiGRU network is good at processing, the sliding time window segmentation method is used to process the normalized data. In this way, the time correlation between adjacent sequences can be fully retained, and the number of training samples can be increased so that the model has higher robustness and generalization ability. For the given time series \( x^j_i \) of system \( i \), where \( F \) and \( T \) are the feature and the time dimension, the sliding window with window width \( S \) slides along its time direction. By super-imposing the time series intercepted by each sliding step length to the third dimension, the tensor with \( (T - S, S, N) \) dimension is formed, and the calculation process is

\[
X_{1:T-S} = x_{1:1+S} \oplus x_{2:2+S} \oplus \cdots x_{i:i+S} \cdots \oplus x_{T-S:T} \tag{7}
\]

where \( X_{1:T-S} \) is the converted 3D tensor, \( x_{i:i+S} \) is a sequence of length \( S \) from \( i^{th} \) time; \( \oplus \) represents a connection of the segmented data in a third dimension, thus forming a three-dimensional tensor.

B. FRAMEWORK OF THE PROPOSED MODEL

Instead of predicting the single-objective RUL value \( y \), the FA-LBiGRU model can predict the LN parameters \( (\mu^j_i, \sigma^j_i) \) that construct PIs of system \( i \) at time \( t \). Exceptionally, by constructing a novel maximum likelihood loss function, LN subject to a couple of parameters \( (\mu^j_i, \sigma^j_i) \) can maximize the probability when \( RUL^j_t = y \). Fig. 7 shows the framework of the proposed innovation model in this manuscript. It consists of the following layers:

1) INPUT LAYER

It can input the historical CM data set to further operation, which is preprocessed and integrated as a 3D tensor with shape \( \text{number of samples} (S), \text{the width of a time window} (T), \text{and number of features} (F) \).

For a matrix sample \( X = (x^1, x^2, \cdots, x^F) = (x_1, x_2, \cdots, x_T)^T \), the expansion can be expressed as

\[
X = \begin{bmatrix} x^1_1 & x^2_1 & \cdots & x^F_1 \\ x^1_2 & x^2_2 & \cdots & x^F_2 \\ \vdots & \vdots & \ddots & \vdots \\ x^1_T & x^2_T & \cdots & x^F_T \end{bmatrix} \in \mathbb{R}^{T \times F} \tag{8}
\]

where \( x_t = [x^1_t, x^2_t, \cdots, x^F_t] (1 \leq t \leq T) \) is the feature vector of length \( F \) at time \( t \), \( x^f = [x^1_f, x^2_f, \cdots, x^F_f] (1 \leq f \leq N) \) is the \( f^{th} \) feature vector of time steps \( T \). Note that \( X \) in Fig. 7 is
the $S^h$th sample of the input data set segmented by the sliding window.

2) FEATURE WEIGHTING LAYER
It is responsible for extracting the attention weight $e$ of input matrix samples $X$, so as to obtain the attention weighted features $X'$. Concretely, the correlation between input data and target modelling parameters is automatically extracted by FA mechanism. Through continuous iteration, the attention weights of each feature at all times are optimized, and then the critical features with enormous contributions to RUL prediction are gradually extracted. According to (1) and (2), attention weight can be expressed as

$$
e = \sigma(W_2\delta(W_1\frac{1}{T}\sum_{t=1}^{T}x_t + b_1) + b_2) \quad (9)$$

Note that $e = [e_1, e_2, \cdots, e_F] \in \mathbb{R}^{F \times 1}$.

Therefore, the attention weighted features are

$$X' = e \odot X = [e_1x_1, e_2x_2, \cdots, e_Fx_F] \quad (10)$$

where $X'$ can record as $[x_1^F, x_2^F, \cdots, x_T^F, \cdots, x_T^F]^{T}$ given $F$ is the dimension of attention-weighted feature, and $T$ is the width of time window.

3) FEATURE LEARNING LAYER
For the system RUL prediction task, the input sensor data feature is a continuous-time series with solid correlation, which requires the network to have a particular “memory function.” The RUL of the current device is predicted by learning the difference before and after degradation. Feature weighted layer can give different weights to multidimensional features, but it is challenging to learn sequence information with obvious time correlation. Therefore, the BiGRU network is built after the weighted feature layer to form the feature learning layer so that the prediction model can capture the temporal dependence of data.

Since the length of time series is $T$, and there are $F$ input features, a BiGRU layer is composed of $T$ memory units in series, and the input dimension of each GRU unit is $F$. The feature sequence is input to the forward and reverse layers at different times, as shown in Fig. 5, to get the forward output $h_f$ and backward output $h_b$ at time $t$, $t \in [1, T]$. Finally, the output $h$ of the unit is calculated by (5), and the total output $H$ of BiGRU can be denoted as $[h_1, h_2, \cdots, h_t, \cdots, h_T]$.

4) OUTPUT LAYER
A novel layer called lognormal layer, which forms an output layer together with Linear, is defined to seek RUL distribution parameters for specific features. The output of the feature learning layer is used as the input of the output layer, and then the PDF parameter prediction values $y^*_{i}$ of the system RUL distribution are calculated, which construct PIs at time $t$, given by

$$y^*_{i} = [\mu_t, \sigma_t] = w_o h_t + b_o \quad (11)$$

where, $w_o$ and $b_o$ are the weight matrix and bias vector of the Linear.

However, the practical construction of PIs requires a slight positive standard deviation, while the ordinary activation function cannot meet this requirement. Hence, new standard
deviation $\sigma'_i$ can be obtained as follows:

$$\sigma'_i = \ln(1 + e^{\sigma_i})$$

(12)

Then note that the total output $Y^*$ of the output layer can be recorded as

$$Y^* = \left[\mu_1, \mu_2, \cdots, \mu_T, \cdots, \mu_T, \sigma'_1, \sigma'_2, \cdots, \sigma'_i, \cdots, \sigma'_T\right]^T$$

(13)

5) LOSS FUNCTION

So far, the FA-BiGRU model without log-norm distribution has been built to construct RUL PIs, determined by sets of parameters $(\mu'_i, \sigma'_i)$ for system $i$ at time $t$, $i \in [1, N]$, $t \in [1, T]$. When training, the model whose name modifies to FA-LBiGRU updates weights and biases with LN participation, which will be altered to get parameters’ sets $(\mu'_i, \sigma'_i)$ that maximize the likelihood function. In other words, according to the probability density formula in Table 1, all the probability values from the start to the end of the monitoring are recorded so that the two parameters tend to the actual RUL distribution, given as

$$L(\mu'_i, \sigma'_i|Y^{(0,2)}_i) = \prod_{j=0}^{t} \frac{1}{Y'_i \sigma'_i \sqrt{2\pi}} e^{-\frac{(\ln Y'_i - \mu'_i)^2}{2(\sigma'_i)^2}}$$

(14)

where $Y^{(0,2)}_i = \{Y^0_i, Y^1_i, \cdots, Y'_i, \cdots, Y'_T\}$.

Whereas, in the actual training process, when the parameters are randomized away from their target values, the likelihood function is relatively constant, and the network update will be minor. Therefore, the logarithmic likelihood function is often used to accelerate the convergence of the model to the optimal value and can calculate the sum of single independent probability factors, which is helpful to the reverse derivation of the loss function. Besides, deep learning methods usually minimize the loss function, so the loss function is defined as

$$E_L = -\sum_{i=1}^{N} \sum_{t=1}^{T_f} \ln L(\mu'_i, \sigma'_i|Y^{(0,2)}_i)$$

(15)

IV. EXPERIMENTAL VERIFICATION

In this section, a well-known prognostic benchmarking problem, i.e. the turbofan engine degradation problem, will be settled for RUL prediction and uncertainty quantification of systems using developed prognostic approaches. Section A presents the benchmarking dataset selected in this article and briefly introduces the pretreatment process. Then performance metrics for point-wise and interval prediction are demonstrated in section B. Section C is dedicated to analyzing the model hyperparameter setting process and results. Visualization and comparative analysis of prognostic results for complex system are described in section D.

A. DATASET DESCRIPTION AND PREPROCESSING

The engine degradation dataset C-MAPSS for prognostic is published by [30], which consists of four subsets (named FD001 $\sim$ FD004), and uses 21 sensors as typical features to characterize engine performance. Each subset comprises of a training set, a test set, and a RUL label set. The training set contains all the data for the whole life cycle of the engine, whereas the test set only contains the data from the previous period of the engine’s life cycle. The RUL label is the RUL of the final monitoring time of the engine corresponding to the test set. Moreover, each set (see Table 2 for specific information) contains a different number of engines, each with varying degrees of initial wear, so the sequence length of each engine monitoring data is also different.

**TABLE 2. Basic information of CMAPSS datasets.**

| Dataset description | C-MAPSS subsets |
|---------------------|-----------------|
|                     | FD001 | FD002 | FD003 | FD004 |
| Operative conditions| 1     | 6     | 1     | 6     |
| Failure modes       | 1     | 1     | 2     | 2     |
| Train engine units  | 100   | 260   | 100   | 249   |
| Test engine units   | 100   | 259   | 100   | 248   |
| Train data volume   | 20631 | 53759 | 24720 | 61249 |
| Test data volume    | 13096 | 33991 | 16596 | 41214 |
| Train minimum lifespan(cycles) | 128  | 128  | 145  | 128  |
| Train maximum lifespan(cycles) | 362  | 378  | 525  | 543  |
| Minimum test length(cycles) | 31   | 21   | 38   | 19   |
| Maximum test length(cycles) | 303  | 367  | 475  | 486  |

In order to deal with complex and massive datasets effectively, a preprocessing method is adopted, which is mainly divided into four steps corresponding to A in section III. In particular, this section mainly describes the process of selecting sensor data, while data normalization, sliding window pretreatment and RUL label rectification need not be introduced because of their broad applicability. Due to space constraints, only the distribution of four sensor data features with different trends in the FD001 training set is shown in Fig. 8. It can be seen that the trend of T2 remains unchanged, indicating that the monitoring data of the No2 sensor have no significant changes in the whole life cycle. Therefore, these sensors with no change are difficult to provide any degradation information that contributes to RUL prediction and should be eliminated. In addition, Table 3 shows trendability, monotonicity and prognosability parameters to compare whether remaining candidate sensors are helpful for system-based prognosis. The most suitable sensors whose parameter is close to “1” can be determined by giving a sum function of the three parameters to compare, which can provide helpful degradation information. Finally, after comprehensively measuring all parameters and their sum in Table 3, ten sensors are selected as the input feature data set in the manuscript: No. 2, 3, 4, 7, 11, 12, 15, 17, 20 and 21, respectively.

B. PERFORMANCE EVALUATION METRICS

Performance evaluation metrics to quantify the performance of prediction model is divided into two aspects: point-wise prediction and uncertainty quantification. Among them, the point-wise prediction evaluation metrics is used to evaluate
FIGURE 8. No.1, No.3, No.9, and No.12 sensor tendency; the dark blue point represents the degradation data, the red point represents the final failure point, the abscissa is the engine’s running time in the whole life cycle, and the ordinate with different units is the measured value of different sensors.

TABLE 3. Parameter suitability metrics of candidate sensors.

| Sensor | Symbol | Monotonicity | Prognosability | Trendability | Sum |
|--------|--------|--------------|----------------|--------------|-----|
| 2      | T24    | 0.811        | 0.822          | 0.842        | 2.475 |
| 3      | T30    | 0.798        | 0.814          | 0.824        | 2.436 |
| 4      | T350   | 0.863        | 0.867          | 0.914        | 2.644 |
| 7      | P30    | 0.782        | 0.742          | 0.703        | 2.227 |
| 8      | Nf     | 0.498        | 0.422          | 0            | 0.92  |
| 9      | Nc     | 0.552        | 0.315          | 0            | 0.867 |
| 11     | Ps30   | 0.904        | 0.893          | 0.926        | 2.723 |
| 12     | Phi    | 0.791        | 0.785          | 0.713        | 2.289 |
| 13     | NFm    | 0.478        | 0.424          | 0            | 0.902 |
| 14     | NRc    | 0.581        | 0.284          | 0            | 0.865 |
| 15     | BPR    | 0.857        | 0.814          | 0.872        | 2.543 |
| 17     | intBleed | 0.828   | 0.83           | 0.841        | 2.499 |
| 20     | W31    | 0.739        | 0.765          | 0.702        | 2.206 |
| 21     | W32    | 0.744        | 0.766          | 0.747        | 2.257 |

1) POINT-WISE PREDICTION EVALUATION METRICS
Given $m$ engines, where $y_i$ is the actual RUL value of the $i^{th}$ engine, while $\hat{y}_i$ is its predicted value, the prognostic model’s performance can be quantified by the following metrics.

a: ROOT MEAN SQUARE ERROR (RMSE)  
RMSE is used to measure the deviation between the predicted value and the actual value, with its calculation function as

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2}$$ (16)

b: SCORING FUNCTION (SF)  
Due to the high safety requirements of the engine, the cost caused by the catastrophic consequences of not timely maintenance is much larger than that of excessive maintenance, so earlier predictions are more desirable than delayed predictions. SF, as an authoritative official indicator in the 2008 PHM data challenge, imposes higher penalties for RUL delayed forecasts, and its calculation formula follows as

$$Score = \left\{ \begin{array}{ll}
\sum_{i=1}^{m} e^{-(y_i - \hat{y}_i)} - 1\hat{y}_i - y_i < 0 \\
\sum_{i=1}^{m} e^{(y_i - \hat{y}_i)} - 1\hat{y}_i - y_i \geq 0
\end{array} \right.$$ (17)

where $a_1$ and $a_2$ are parameters controlling the degree of punishment of SF for prediction lag to achieve asymmetric prediction preference, which equals 13 and 10 respectively.

2) UNCERTAINTY QUANTIFICATION EVALUATION METRICS OF COMPLEX SYSTEM
a: PREDICT INTERVAL COVERAGE PROBABILITY (PICP)  
PICP, which can be used to evaluate the reliability of PIs, reflects the probability that the actual RUL value $y_i$ falls into the predicted fluctuation range $[L_\alpha(\hat{y}_i), U_\alpha(\hat{y}_i)]$ of the predicted value $\hat{y}_i$.  

$$PICP = \frac{1}{m} \sum_{i=1}^{m} \kappa(y_i); \kappa(y_i) = \left\{ \begin{array}{ll}
1, & \text{if } y_i \in [L_\alpha(\hat{y}_i), U_\alpha(\hat{y}_i)] \\
0, & \text{if } y_i \notin [L_\alpha(\hat{y}_i), U_\alpha(\hat{y}_i)]
\end{array} \right.$$ (18)

where $L_\alpha(\hat{y}_i)$ and $U_\alpha(\hat{y}_i)$ are the upper and lower boundaries of the PI of $\hat{y}_i$ at the nominal confidence level (CL) $(1 - \alpha)$, respectively. The larger the PICP is, the more real values fall into the PI, and the more reliable the prediction results of the model are. In addition, PICP should exceed the given confidence as much as possible. Otherwise, it will be judged invalid in the actual maintenance decision.

b: PREDICT INTERVAL NORMALIZED AVERAGE WIDTH (PINAW)  
The pursuit of PICP alone will lead to too broad a PI, and it is difficult to obtain the uncertainty information of RUL.
prediction accurately. Therefore, it is necessary to consider the interval width to improve the decision value. PINAW represents the average width of the PI, which can be used to assess the clarity of the PI.

\[
PINAW = \frac{1}{m(y_{\text{max}} - y_{\text{min}})} \sum_{i=1}^{m} (U_\alpha(\hat{y}_i) - L_\alpha(\hat{y}_i)) \quad (19)
\]

where \(y_{\text{max}}\) and \(y_{\text{min}}\) are the maximum and minimum of the real RUL, respectively. When PICP is constant, the smaller PINAW corresponds to the narrower PI, the better the prediction effect of the model.

C. MODEL HYPER-PARAMETER SETTING

In order to accurately reflect the error between the predicted value and the actual value of different hyper-parameter models in the test set, the commonly used mean square error (MSE) is used as the evaluation metrics to adjust the hyper-parameter.

\[
\text{MSE} = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \quad (20)
\]

where \(y_i\) and \(\hat{y}_i\) are the actual and predicted values of the \(i^{th}\) input sample, respectively; \(m\) is the number of samples.

This model is implemented by the PyTorch framework, and the main hyper-parameters involved are: network structure, learning rate, batch size, epoch, optimizer, dropout rate, etc. The model parameters have a great influence on the performance of the model, especially the time window width \(S\) and batch size. Therefore, the trial-and-error method is used to minimize the model’s prediction error to obtain the optimal super-parameter combination, as shown in Table 4. Concretely, taking the above two parameters as an example, Fig. 9 shows the specific adjustment.

MScore and MMSE are the mean values of 30 test results in Fig. 9, respectively, and Score is the SF introduced in B. Considering that the time window width should not exceed the minimum length of engine monitoring in the dataset, its test range is \([5, 30]\), and the test range of batch size is \([8, 64]\). In addition, more useful degradation information that may continuously enhance model’s prognostic performance is accumulated due to time window width increase. Finally, comparing the lowest MScore and MMSE with the best performance, a time window width of 30 cycles and batch size of 32 are selected as the most reasonable parameters of the model, and other hyper-parameters are adjusted according to this method.

D. PROGNOSTIC RESULTS OF COMPLEX SYSTEM

1) VISUALIZATION ANALYSIS OF RUL PREDICTION AND UNCERTAINTY ESTIMATION

Taking the FD001 dataset as an example, the preprocessed high-dimensional time series was loaded into the FA-LBiGRU model for the point prediction of RUL, and the median prediction results of the test set are given in Fig. 10-a. Furthermore, all the engines in the FD001 test set are reordered from large to small according to the real RUL values, and the results are shown in Fig. 10-b.

It clearly can be seen that the actual value and the median value of predicted RUL value are very close in Fig 10-a, indicating the effectiveness of the point-wise prediction of the proposed method. Additionally, when the RUL value is considerable, the engine is healthy, indicating that the system operates well. Still, the monitoring data are insufficient to obtain sufficient degradation information. Therefore, as shown in Fig. 10-b, when the engine starts running, the RUL prediction error is large, and the fluctuation is relatively violent and shows noticeable hysteresis. However, after a long time of operation, the predicted RUL converges around the real RUL, and the prediction performance is significantly enhanced. Thus, it can be concluded that the more sufficient the history information of the engine, the more pronounced the performance degradation information, and the smaller the model’s prediction error.

Next, Fig. 11 and Fig. 12 respectively present the RUL prognostics and interval construction results of engine #34 and engine #76 in the test subset of FD001. Fig. 11-a and Fig. 12-a show the predicted PDFs and RULs at the last period of all monitoring times respectively. It can be observed that the median value of predicted RUL (characterized by the red star point in Fig. 11-a and Fig. 12-a is close to the actual
value (characterized by the black hollow circle in Fig. 11-a and Fig. 12-a) at the end of the operation (155-175 monitoring times). Subsequently, the peak value of PDF at each time point increases with monitoring time, demonstrating that the point-wise prediction results of RUL are convincing in developing maintenance strategies.

Fig. 11-b and Fig. 12-b respectively illustrate the RUL prognostics and interval construction results throughout
TABLE 5. Comparison of the point prediction results using different models.

| Method          | FD001       | FD004       |
|-----------------|-------------|-------------|
|                 | RMSE        | SF          | Training time(s) | Testing time(s) | RMSE        | SF          | Training time(s) | Testing time(s) |
| RNN             | 15.96       | 382         | 261             | 1.2             | 26.54       | 6592        | 784             | 2.5             |
| LSTM            | 14.92       | 342         | 297             | 1.3             | 25.16       | 5127        | 916             | 2.4             |
| GRU             | 15.14       | 356         | 265             | 0.8             | 26.17       | 5550        | 803             | 1.9             |
| BiGRU           | 14.53       | 321         | 322             | 1.0             | 24.26       | 4630        | 965             | 2.2             |
| SA+BiGRU        | 13.86       | 291         | 343             | 1.2             | 23.45       | 4886        | 983             | 2.5             |
| IESGP[23]       | 14.72       | 332         | 365             | 1.5             | 28.61       | 6281        | 1096            | 2.9             |
| LSTMBS[26]      | 14.46       | 481         | 334             | 1.2             | 27.11       | 5200        | 971             | 2.6             |
| FA-LBiGRU Median | 13.13       | 238         |                 | 1.3             | 20.29       | 3395        |                 |                 |
| FA-LBiGRU Mode  | **12.76**   | 249         | 349             |                 | **19.75**   | 3442        | 996             | 2.6             |
| FA-LBiGRU Mode  | 13.75       | 254         |                 |                 | 21.54       | 3484        |                 |                 |

TABLE 6. Detailed settings for comparison methods without citations.

| Hyper-parameters | RNN LSTM GRU BiGRU SA+BiGRU |
|------------------|-----------------------------|
| optimizer        | Adam                        |
| epoch            | 150                         |
| dropout rate     | 0.25                        |
| learning rate    | 0.004 0.004 0.002 0.002 0.002 |
| batch size       | 64 64 32 32 32              |
| Neurons of model | 256 256 128 128 128         |
| Layers of model  | 6 6 8 8 8                  |
| Neurons of SA    | N/A N/A N/A N/A 24         |

 lifetime observations of engine #34 and engine #76 in the test set FD001. For lifetime-cycle observations, as the degradation process is not noticeable compared to the end-of-life observations, the predicted RUL distribution in the early phase of the observation period is unsatisfied: ground actual RUL may exceed the 95% RUL prediction CL, e.g. about 55 cycles in Figure 12-b. However, as shown in Fig. 11-b and Fig. 12-b, the prediction results are improved along the engine’s lifetime: the widths of the 95%, 90% and 80% CL intervals become narrower as the monitored system runs, which means that the uncertainty of prediction decrease gradually with the number of collected measurements increasing. It also can be seen that the predicted RUL is getting closer to the ground actual one as the monitored system runs, which indicates that the performance degradation of the engine gradually accumulates with the increase of the operation cycle, making the model capture the degradation information in the time series, which effectively improves the prediction accuracy in the later stage of operation. In fact, at the end of the engine #34 observation period (cycle 173), its actual RUL (7 cycles) falls within a narrow enough 80% CL RUL interval [4.0, 7.3] (Fig. 11-c). Moreover, at the end of the engine #76 observation period (cycle 175), its actual RUL (10 cycles) falls within a pretty narrow interval [9.7, 15.8], representing the RUL PI of 80% CL (Fig. 12-c).

2) COMPARATIVE ANALYSIS WITH STATE-OF-THE-ART METHODS

To compare the last prediction performance of this method for engine RUL point prediction and interval prediction, the metrics in Section B are used to objectively evaluate the accuracy and generalization ability of different models on the test set. Table 5 compares the training effects of different recurrent neural networks, using the FD001 dataset in a single mode and the FD004 dataset in multiple modes. Notably, the feature extraction effects of different attention mechanisms, the self-attention (SA) and the FA mechanism, are compared. Detailed settings for comparison methods without citations are shown in Table 6. Furthermore, the point-wise prediction accuracy of the IESGP and LSTMBS, which can also be used for interval prediction, is also compared. In addition, observing the results in Table 5, owing to the asymmetric property of the LN, there exists a subtle disparity when using the mode, mean or median values for the evaluation of the point-wise prediction accuracy metrics. Among them, the mean provides better results for the RMSE than the other two accuracy metrics. Nevertheless, comprehensively considering the disaster of lag prediction, the median with a smaller SF value is selected as the result of the point prediction.

Compared with the other state-of-the-art methods according to Table 5, the RMSE and SF of proposed method are minimal, denoting that the proposed FA-LBiGRU in this paper has the highest prediction accuracy. For the FD001 dataset, RMSE = 13.13, where is reduced by an average of 11% compared with other methods, and SF = 238, where is reduced by an average of 33%. For the FD004 dataset, RMSE = 20.29, where is reduced by an average of 22% compared with other methods, and SF = 3395, where is reduced by an average of 37%. Further analysis shows that although GRU training is slightly less effective than LSTM, its training time is reduced by about 11%, which is conducive to real-time prediction deployment. Secondly, BiGRU can capture bidirectional long-time-dependent features, resulting in better prediction accuracy. Moreover, the FA mechanism
can score the importance of different features rather than the weighted normalization of the SA mechanism, which further improves the model’s prediction accuracy. In addition, although the average time-consuming of the proposed method in the offline training phase is 349s, the online test phase only takes 1.3s, which is only a 0.13s difference compared with the average time-consuming by other methods applied to FD001. On the other hand, the point prediction accuracy of the proposed method is higher, which can compensate for the time loss caused by the model’s complexity. In conclusion, the longitudinal comparison of the point prediction results shows that the proposed method has higher prediction accuracy and better generalization ability while meeting the real-time requirements.

To further verify the advantages of the proposed method for interval prediction, IESGP [23], the BDL-based [24], and LSTMBS [26] are constructed as comparison models. The average interval predictions are obtained by 10 trials on the FD001 dataset, as shown in Figure 13. Compared to other methods, the proposed model provides the tightest width of 95% CL RUL prediction interval (PINAW = 0.3) with a good enough score of the prediction interval coverage percentage (PICP = 0.95). It can be seen that IESGP gets the best PICP, but its mean prediction interval width is more than a half of the target RUL’s range (PINAW = 0.540), indicating that it is slightly weak when facing high-dimensional features. Besides, although the PICP of LSTMBS has the same value as our model, the average PINAW decreases by 19.8%, 13.9%, and 16.9% at the CL of 95%, 90%, and 80%, respectively. In addition, the repeated test error of the proposed method is relatively low, indicating that the proposed method can quantify the uncertainty of performance degradation more accurately, and stably and effectively improve the prediction accuracy. The accuracy of point prediction is also better than the LSTMBS model, indicating that the proposed method has a more powerful feature extraction ability and data processing power.

To sum up, it is found that the proposed method can extract more accurate performance degradation trend characteristics from high-dimensional and sparse big data and effectively measure the uncertainty in the process of system degradation to achieve accurate prediction of RUL. The prediction interval has the advantages of high reliability and accuracy, which provides strong support for subsequent operation and maintenance personnel to make maintenance decisions.

V. CONCLUSION

In this work, a novel end-to-end methodology for RUL estimation and uncertainty quantification of a system was presented. The multidimensional original monitoring data is pre-processed into a data set for model input by data preprocessing technology. Additionally, the degradation rule of data sets is learned to predict the RUL distribution of complex system by FA-LBiGRU model. The famous benchmark data set C-MAPSS is used to verify the effectiveness and advantages of this method, and the following conclusions are yielded:

1) Sufficient historical data of systems contain obvious performance degradation information, which reduces the error between point prediction and interval prediction of the model. It is conducive to health management measures such as subsequent operation planning and maintenance decision-making.

2) FA mechanism makes end-to-end prognostics models focus more on essential features, and dynamic weighting of the importance of multidimensional features can be achieved without manual intervention. LBiGRU can deeply mine the degradation information before and after time series and measure the uncertainty of equipment degradation process based on LN distribution, effectively improving the accuracy and reliability of remaining life point-wise prognostics and interval prognostics of mechanical system.

3) Compared with several state-of-the-art approaches, the FA-LBiGRU fusion model is used to process the monitoring data of multi-dimensional mechanical systems with complex operating conditions and variable fault modes, which can generate the RUL PIs taking reliability and accuracy into account.

REFERENCES

[1] F.-K. Wang, X.-B. Cheng, and K.-C. Hsiao, “Stacked long short-term memory model for proton exchange membrane fuel cell systems degradation,” J. Power Sources, vol. 448, Feb. 2020, Art. no. 227591.

[2] Y. Li, X. Huang, C. Zhao, and P. Ding, “A novel remaining useful life prediction method based on multi-support vector regression fusion and adaptive weight updating,” ISA Trans., Apr. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S001905782200204X
[1] X. Xu, X. Li, W. Ming, and M. Chen, “A novel multi-scale CNN and attention mechanism method with multi-sensor signal for remaining useful life prediction,” Comput. Ind. Eng., vol. 169, Jul. 2022, Art. no. 108204.

[2] S. V. Angadi and R. L. Jackson, “A critical review on the solenoid valve reliability, performance and remaining useful life including its industrial applications,” Eng. Failure Anal., vol. 136, Jun. 2022, Art. no. 106231.

[3] 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.

[4] L. Yan, J. Peng, D. Gao, Y. Wu, Y. Liu, H. Li, W. Liu, and Z. Huang, “A hybrid method with cascaded structure for early-stage remaining useful life prediction of lithium-ion battery,” Energy, vol. 243, Mar. 2022, Art. no. 128338. [Online]. Available: https://ideas.repec.org/a/eee/energy/v243y2022is03605442210532874.html

[5] L. Liu, X. Song, and Z. Zhou, “Aircraft engine remaining useful life estimation via a double attention-based data-driven architecture,” Rel. Eng. Syst. Saf., vol. 221, May 2022, Art. no. 108330.

[6] J. Wu, K. Hu, Y. Cheng, H. Zhu, X. Shao, and Y. Wang, “Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network,” ISA Trans., vol. 97, pp. 241–250, Feb. 2020.

[7] N. Li, P. Xu, Y. Lei, X. Cai, and D. Kong, “A self-data-driven method for remaining useful life prediction of wind turbines considering continuously varying speeds,” Mech. Syst. Signal Process., vol. 165, Feb. 2022, Art. no. 108315.

[8] C. Chen, J. Shi, N. Lu, Z. H. Zhu, and B. Jiang, “Data-driven predictive maintenance strategy considering the uncertainty in remaining useful life prediction,” Neurocomputing, vol. 494, pp. 79–88, Jul. 2022.

[9] T. Mamno and F.-K. Wang, “Long-term short-term memory with attention mechanism for state of charge estimation of lithium-ion batteries,” IEEE Access, vol. 8, pp. 94140–94151, 2020.

[10] X. He, Z. Wang, Y. Li, S. Khazhina, W. Du, J. Wang, and W. Wang, “Joint decision-making of parallel machine scheduling restricted in job machine release time and preventive maintenance with remaining useful life constraints,” Rel. Eng. Syst. Saf., vol. 222, Jun. 2022, Art. no. 108429.

[11] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, “Long-short term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries,” IEEE Trans. Veh. Technol., vol. 67, no. 7, pp. 5695–5705, Jul. 2018.

[12] Y. Cao, M. Jia, P. Ding, and Y. Ding, “Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network,” Measurement, vol. 178, Jun. 2021, Art. no. 109287.

[13] Q. Wu, K. Ding, and B. Huang, “Approach for fault prognosis using recurrent neural network,” J. Intell. Manuf., vol. 31, no. 7, pp. 1621–1633, Oct. 2020.

[14] Y. Li, X. Huang, P. Ding, and C. Zhao, “Wiener-based remaining useful life prediction of rolling bearings using improved Kalman filtering and adaptive modification,” Measurement, vol. 182, Sep. 2021, Art. no. 109706.

[15] J. Lin, G. Liao, M. Chen, and H. Yin, “Two-phase degradation modeling and remaining useful life prediction using nonlinear Wiener process,” Comput. Ind. Eng., vol. 160, Oct. 2021, Art. no. 107533.

[16] K. Song and L. Cui, “A common random effect induced bivariate gamma degradation process with application to remaining useful life prediction,” Rel. Eng. Syst. Saf., vol. 219, Mar. 2022, Art. no. 108200.

[17] Z. Lyu, R. Gao, and L. Chen, “Li-ion battery state of health estimation and remaining useful life prediction through a model-data-fusion method,” IEEE Trans. Power Electron., vol. 36, no. 6, pp. 6228–6240, Jun. 2021.

[18] N. Li, N. Gebraeel, Y. Lei, X. Fang, X. Cai, and T. Yan, “Remaining useful life prediction based on a multi-sensor data fusion model,” Rel. Eng. Syst. Saf., vol. 208, Apr. 2021, Art. no. 107249.

[19] P. Kundu, A. K. Darpe, and M. S. Kulkarni, “Weibull accelerated failure time regression model for remaining useful life prediction of bearing working under multiple operating conditions,” Mech. Syst. Signal Process., vol. 134, Dec. 2019, Art. no. 106302.

[20] X. Wu, K. Ding, J. Dong, and Y. Pi, “A deep belief network based health indicator construction and remaining useful life prediction using improved particle filter,” Neurocomputing, vol. 361, pp. 19–28, Oct. 2019.

[21] C. Liu, L. Zhang, Y. Liao, C. Wu, and G. Peng, “Multiple sensors based prognostics with prediction interval optimization via echo state Gaussian process,” IEEE Access, vol. 7, pp. 112397–112409, 2019.

[22] C.-H. Hu, H. Pei, X.-S. Si, D.-B. Du, Z.-N. Pang, and X. Wang, “A prognostic model based on DBN and diffusion process for degrading bearing,” IEEE Trans. Ind. Electron., vol. 67, no. 10, pp. 8767–8777, Oct. 2020.

[23] Y. Liao, L. Zhang, and C. Liu, “Uncertainty prediction of remaining useful life using long short-term memory network based on bootstrap method,” in Proc. IEEE Int. Conf. Prognostics Health Manag. (ICPHM), Jun. 2018, pp. 1–8.

[24] W. Peng, Z.-S. Ye, and N. Chen, “Bayesian deep-learning-based health prognostics toward prognostics uncertainty,” IEEE Trans. Ind. Electron., vol. 67, no. 3, pp. 2283–2293, Mar. 2020.

[25] O. Bektas, J. A. Jones, S. Sankararaman, I. Roychoudhury, and K. Goebel, “A neural network filtering approach for similarity-based remaining useful life estimation,” Int. J. Adv. Manuf. Technol., vol. 101, nos. 1–4, pp. 87–103, Mar. 2019.

[26] Z. Que, X. Jin, and Z. Xu, “Remaining useful life prediction for bearings based on a gated recurrent unit,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–11, 2021.

[27] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, “Long-short term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries,” IEEE Trans. Veh. Technol., vol. 67, no. 3, pp. 6228–6240, Jun. 2021.

XIAOJIA YAN received the B.S. degree from the School of Mechanical and Electrical Engineering, Wuhan University of Technology, Wuhan, China, in 2020. He is currently pursuing the M.S. degree in weapon science and technology with the Naval University of Engineering (NUE), Wuhan. His current research interests include deep learning, signal processing, and remain useful life prediction of complex systems.

WEIGE LIANG received the B.S. and Ph.D. degrees in weapon science and technology from the Naval University of Engineering (NUE), Wuhan, China, in 2005 and 2010, respectively. He is currently an Instructor with the School of Weapon Engineering, NUE. His research interests include machine learning, prognostic health management, and control theories and applications.

DONGXUE XU received the B.S. degree from the School of Mechanical Engineering, Huazhong University of Science and Technology (HUST), China, in 2015. He is currently pursuing the M.S. degree with the Naval University of Engineering, China. His research interests include signal processing, deep learning, guidance, and control.