

Received May 14, 2020, accepted May 23, 2020, date of publication May 27, 2020, date of current version June 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2998003

# Remaining Useful Performance Estimation for Complex Analog Circuit Based on Maximal Information Coefficient and Bidirectional Gate Recurrent Unit

BOLUN DU<sup>1</sup>, YIGANG HE<sup>1</sup>, (Member, IEEE), BAORAN AN<sup>2</sup>, AND CHAOLONG ZHANG<sup>1</sup>

<sup>1</sup>School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China

<sup>2</sup>Institute of Computer Application, China Academy of Engineering Physics, Mianyang 621000, China

Corresponding author: Yigang He (yghe1221@whu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51977153, Grant 51977161, Grant 51607004, and Grant 51577046, in part by the State Key Program of National Natural Science Foundation of China under Grant 51637004, in part by the National Key Research and Development Plan Important Scientific Instruments and Equipment Development under Grant 2016YFF0102200, and in part by the Equipment Research Project in Advance under Grant 41402040301.

**ABSTRACT** Degradation of circuit components are typically accompanied by a deviation in component parameters from their normal values, which can ultimately influence the stable operation of complex analog circuit. To address this concern, remaining useful performance (RUP), regarded as the useful performance from the current time to the end of performance, is an effective way to ensure system safety by providing early warning of failure and enabling forecast maintenance. In this paper, a novel RUP estimation method based on the two-stage maximal information coefficient (TSMIC) and bidirectional gate recurrent unit (Bi-GRU) network is proposed. Initially, the run to failure data of the circuit in real-time is obtained by RT-LAB hardware-in-the-loop. Additionally, to obtain suitable features reflecting degradation trend over cycles, a TSMIC method is proposed to eliminate features hardly changing with degradation cycle in the first stage, mine mutual information between features in the second stage. Furthermore, the linear regression model is used as a performance evaluation to retain the original pattern in the selected features. Through the fusion of the selected multi-features, health indicators of different circuit components are constructed. Ultimately, the deep Bi-GRU unit network, which can extract representative time-series information and explore subtle differences of the degradation cycles, is used to generate prediction results. The proposed framework is verified through a case study on the complex analog circuit, and comparisons with other state-of-the-art methods are presented. The experimental results of the case study show the effectiveness and superiority of the proposed approach.

**INDEX TERMS** Features selection, maximal information coefficient, recurrent neural network, remaining useful performance.

## I. INTRODUCTION

Inertial confinement fusion (ICF) [1] is one of the fascinating approaches to achieve fusion ignition, and there have been several huge laser drivers around the world for such research, i.e., the Laser Megajoule in France and National Ignition Facility (NIF) in the United States [2]. The SG-III laser facility is the largest laser driver in China, which is carried out in the Laser Fusion Research Center of China Academy of

Engineering Physics (CAEP). As the most important equipment of the SG-III laser fusion facility, the MJ-level power condition module (PCM) circuit consists of 53 components in total, which is regarded as a complex analog circuit. The circuit components deviate from their nominal values gradually under the condition of high voltage and high current pulse. This gradual degradation will affect the performance of the circuit, even resulting in catastrophic accidents. Therefore, it is of great significance to propose a prognostics and health management (PHM) approach to facilitate early warning and predictive maintenance of circuit failures.

The associate editor coordinating the review of this manuscript and approving it for publication was Baoping Cai<sup>1</sup>.

In the past decades, PHM [3], [4] technology have been extensively studied in various engineered systems, i.e., health assessment for electronic component [5], fault diagnosis for three-phase inverters [6], remaining useful life (RUL) prediction for IGBT [7], subsea pipelines [8] and batteries [9]. PHM for circuit system can be divided into the following: 1) the detection and fault isolation (FDI) [10]–[13] of faults and 2) the prediction of the remaining useful performance (RUP) [14] of the failing circuit. Nowadays, the PHM researches of a circuit are mainly focused on FDI technology, and there are few works of literature on RUP estimation of a circuit. Nevertheless, FDI technology cannot prevent the faults of the circuit, since it is utilized to process the abnormal output information of the analog circuit after the fault occurs. The faults of the circuit are generally caused by the gradual aging of the components, which makes the component values gradually deviate from the normal value, thereby reaching the failure threshold that exceeds the normal tolerance range [15]. Therefore, accurate and robust RUP estimation for the circuit provides a basis for timely and reasonable maintenance to avoid circuit disaster.

The concept of using RUP to predict the circuit system is first put forward by the center for advanced life cycle engineering (CALCE) of Maryland University in 2013 [14]. Here, RUP estimation refers to the prediction of the degradation cycle when the circuit is no longer be able to perform its intended function [16]. RUP estimation is made challenging due to the existence of component tolerances, the complex fault mechanism, and the interdependence of the circuit component. **Section II** comprehensively reviews the state-of-the-art RUP estimation approaches for analog circuit in recent years, analyzing the strengths and limitations of these approaches. Generally, RUP estimation approaches for an analog circuit can be divided into model-based [17] and data-driven approaches [18]. Although these RUP estimation approaches achieve good results for the analog circuit, there are still some deficiencies as follows:

- The existing researches on circuit RUP estimation mainly focus on DC-DC converters [19], and simple analog circuit, e.g., Sallen-Key bandpass and Biquad low-pass filter circuit (refer to excellent literature [14], [16], [17]). Additionally, the existing model-based RUP estimation approaches show difficulties in building precise degradation mechanism models in practice owing to the uncertainty of circuits and sensor noise.
- As for feature selection, the existing methods show ignorance in considering the nonlinear correlations among features, which exists in most circuit systems. As for prognostics algorithm, the shallow learning models, such as support vector regression (SVR) and relevance vector regression (RVR), often suffer from invalid learning and weak generalization when learning with a large amount of feature data.

This paper proposed a novel prognostic method based on the two-stage maximal information coefficient (TSMIC) and bidirectional gate recurrent unit (Bi-GRU) network to deal

with the above shortcomings. The technical contributions are summarized as follows:

- A comprehensive metric, namely, TSMIC, which is proposed to select optimal features with deep nonlinear correlations between features following the overall degradation cycle. In the first stage, the main feature set is obtained, where redundant features having tiny correlations with the degradation cycle are eliminated. In the second stage, the optimal feature subset is acquired by comparing the value of MIC between any two features in the main feature set, which eliminates features that not representative of the degradation trend.
- For RUP estimation, the proposed deep Bi-GRU network is devised for tracking the evolution of the HI and generating the high-quality RUP results for circuit components. It has the advantage of using two hidden layers to process the sequence of feature data in two directions to capture both past and future information, respectively. Compared with different prognosis algorithm, i.e., SVR, deep convolutional neural networks (DCNN), long-short term memory (LSTM), gate recurrent unit (GRU) and bidirectional long short-term memory (Bi-LSTM), the experiment results show that the proposed method achieves relatively higher prognosis accuracy for different circuit components.

The rest of this paper is organized as follows: Section II reviews the related work of RUP estimation methods. Section III introduces the background of the MJ-level PCM circuit. Section IV outlines the proposed RUP estimation framework and theoretical. Section V presents the experimental results of data processing, feature selection, degradation trends construction, and performance prognostics, respectively. Conclusion and future work are drawn in Section VI.

## II. RELATED WORK OF RUP ESTIMATION METHODS

In the following subsections, state-of-the-art model-based and data-driven RUP estimation approaches are reviewed.

### A. MODEL-BASED APPROACHES

Most model-based approaches are dependent on the empirical knowledge of the operation conditions, material characteristics and failure mechanism to build mathematical models, among which, regression-based methods, hidden Markov model [20], Weibull distribution model [21], Kalman filter and Particle filter [22] are representative.

The regression-based method assumes that the future state value of the analog circuit depends linearly on its previous observations and a stochastic term [23]. There are two conventional models: autoregressive and moving average (ARMA) model and autoregressive integrated moving average (ARIMA) model, which have been widely utilized in the analog circuit for RUP estimation. For instance, in [24], the ARMA model predicts the RUP of analog circuits by combining with particle filtering algorithm. In [25], Ibrahim *et al.* employed an ARIMA model and polynomial regression method for RUP estimation in the wavelet domain.

Nevertheless, the regression-based method mainly depends on the trend of historical data of circuit components, which may lead to an inaccurate RUP estimation over the degradation cycle.

Markov models underlying assumptions mean that the degradation processes of the analog circuit should evolve in a finite state-space following the principle of the Markov property [20]. Kharoufeh and Cox [26] proposed a RUP estimation method based on the Markov model and stochastic failure model to calculate the RUP distribution numerically and its moment status. After that, the hidden Markov model has been employed for prognostics to overcome the drawback that the Markov model cannot model the hidden health state of the degradation process. Besides, the Weibull distribution model has been used to study the lifetime distribution of key components of a complex analog circuit such as NIF in the USA [21].

Kalman filter and Particle filter methods are commonly used for RUP estimation of the analog circuit. Kalman filter can update and process the degradation data of circuit components in real-time, and there is no need to store circuit state estimation parameters or circuit observation data. Celaya *et al.* [27] used the accelerated aging test to obtain the degradation trend of the MOSFET on-resistance and then used an extended Kalman filter algorithm to model and predicted the degradation trend of the circuit. In the Particle filter method, Bayes update and particles with probability information of unknown parameters are processed in a certain order, which can solve the time series prediction problem well [22].

In summary, the above model-based approaches can produce accurate prediction results, but the damage mechanics of circuit components must be analyzed. In contrast, data-driven prognostic approaches are attracting increasing attention with their powerful generalization abilities toward complex engineering systems and the rapid development of artificial intelligence.

## B. DATA-DRIVEN APPROACHES

Data-driven approaches mainly include four parts: feature extraction, feature selection, health indicator (HI) or fault indicator (FI) construction, and RUP estimation.

### 1) FEATURE EXTRACTION

The feature sets usually include three types: time-domain features, frequency domain features, and time-frequency domain features. The time-domain features and frequency domain features can capture the amplitude and spectrum differences between fault state and the normal state of the signal, respectively. The time-frequency feature can reflect the joint distribution information in time and frequency domains. In the existing literature, Long *et al.* extracted the high-order time-domain statistics as features for the diagnosis of the analog circuit [28]. The features of the circuit were fetched via fast FFT, sweep frequency response analysis of output signal. The time-frequency analysis, i.e., cross wavelet transform [29], S-transform [30], have been proposed to obtain a

time-frequency signal [31]. In this paper, the output response signal of the MJ-level PCM circuit changes significantly during the whole degradation cycle, and the time-domain feature can reflect more detailed information on the degradation process.

### 2) FEATURE SELECTION

At present, various feature selection metrics have put forward to eliminate redundant features and remain representative features. Javed *et al.* [32] developed a trendability metric, to calculate the relationship between the feature and degradation cycle. Monotonicity was proposed to coincide with the irreversible degradation processes. However, the above metrics are more sensitive to linear correlation than nonlinear correlation. If there is no correlation, weak correlation or redundant correlation in the selected features, the following consequences can be resulted in: 1) When the greater number of features selected, analyze these features will be time-demanding; 2) Too many features selected can cause “dimension disaster”. In contrast, the maximal information coefficient (MIC) is a more effective metric, which always shows a good performance in terms of the ability to detect nonlinear information of data [33]. For example, Bing *et al.* [15] proposed a feature selection algorithm based on MIC to process continuous data and identify the nonlinear function. Zhang *et al.* [16] proposed a McTwo algorithm to measure all the features for the MIC associations with the class labels. Based on these reasons, this paper proposes an improved TSMIC algorithm to obtain the optimal feature set.

### 3) HI/FI CONSTRUCTION

The HI/FI curve characterizes the degradation trend of the circuit. Circuit prognosis is implemented by generating long-term prediction for HI/FI curves until the failure threshold is reached. In the existing literature, in [22], Li *et al.* proposed that FI of the circuit components can be calculated by several special parameter features extracted from the output frequency-domain signals. Yan *et al.* [34] applied the Markov distance to construct a virtual FI from the circuit output response. Zhang *et al.* [16] proposed a continuous parameter calculation method characterized by the correlation of output voltage, which calculated the HI as the  $\cos(*)$  and  $\sin^{-1}(*)$  of the distance between the test feature and the feature extracted from the circuit response. Nevertheless, all the above HI/FI construction methods are idealized, i.e., linear HI/FI curves, without calculating the appropriate HI/FI according to the different degradation effects on the circuit components. Thus, multi-feature fusion approach is proposed in this paper to construct a multi-scale HI model, representing the degradation for MJ-level PCM circuit components.

### 4) PROGNOSTICS ALGORITHM

There are many shallow learning models, i.e., SVR [35] and RVR [36], which have been widely implemented in RUP estimation. Nevertheless, these methods cannot reveal the complex inherent relationships between the root cause of

failure and the signal signatures, which often suffer from invalid learning and weak generalization when learning and training with a large number of fault features.

Nowadays, Bayesian network (BN) [37] or dynamic Bayesian network (DBN) [38] have many applications in fault diagnosis [39] and are increasingly applied in RUL prediction [8]. In the existing literature, Cai *et al.* [8] proposed a hybrid RUL estimation approach of structure systems considering the influence of multiple causes by using DBNs. In [40], this paper used BN models to perform the reliability evaluation of subsea blowout preventer control systems. In addition, DCNNs is emerging as a highly effective neural network architecture for RUP estimation. Lei *et al.* [41] proposed a new data-driven approach based on DCNN by using raw sensor data. Compared to shallower networks, the DCNN has a strong ability to capture basic features, but it ignores the relevance of time series signals over the degradation cycle.

RNNs can learn the latent representation from the entire history of the previous inputs to the target vector by establishing connections between units from a directed cycle. RNN, particularly LSTM first introduced by [42], is devolving as powerful models for reducing the difficulty of learning long-term dependencies in the condition-based maintenance data. In the existing literature, the LSTM based encoder-decoder scheme [43] is proposed to obtain an unsupervised prognosis results using multi-sensor time-series data. However, LSTM can only access previous information instead of future information. Based on the above reasons, in [44], a novel prognostic method based on and Bi-LSTM is developed by Huang *et al.* This method not only can capture the previous information but also can capture future information. In [45], a multi-scale dense GRU network is proposed by Ren *et al.* The GRU chooses a new type of hidden unit that has been motivated by the LSTM unit. It combines with the forget gate and the input gate into a single update gate. It is also mixed with cellular state and hidden state. The final model is simpler than the standard LSTM model and is a very popular variant. Therefore, in this paper, the Bi-GRU are used to process selected features, followed by linear regression layers for generating high-quality prediction results.

### III. BACKGROUND OF SG-III LASER FACILITY

#### A. INTRODUCE OF SG-III LASER FACILITY

As shown in Fig. 1(a) shows the schematic of SG-III laser facility. Fig. 1 (b) and (c) show the architectures and the Beam Reverser in the laser bay. Most architecture has been constructed in the laser bay, including four bundles of laser beams with beam parameter measure assembly, the whole frontend system, energy system, target system and the beam integrated diagnostic system. Therein, the MJ-level PCM circuit is the most important high-power pulsed circuit in energy system. The MJ-level PCM circuit is consist of the pre-ionization and main ionization circuit. Because the circuit system is enormous and cannot be destroyed, we can only

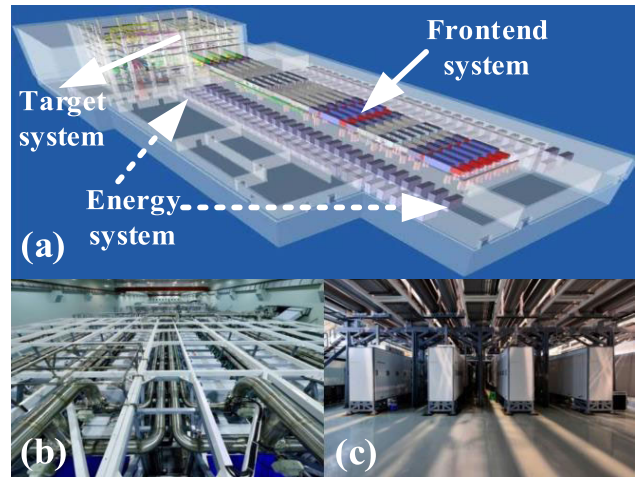


FIGURE 1. (a) Schematic of the SG-III laser facility; (b) and (c) The architectures and the beam reverser in the laser bay (Picture was taken in CAEP).

study the degradation performance of the MJ-level PCM circuit through simulation experiments.

#### B. TOPOLOGICAL GRAPH OF MJ-LEVEL PCM CIRCUIT

Fig. 2. represent the structure of the MJ-level PCM circuit, which includes 53 components in total, i.e., main ionization capacitors ( $C_1 \sim C_{10}$ )/inductors ( $L_1 \sim L_{10}$ )/resistors ( $R_1 \sim R_{10}$ ), pre-ionization capacitor  $C_0$ /inductor  $L_0$ /resistor  $R_0$ , ballast inductors ( $L_{B1} \sim L_{B10}$ ) and Xenon lamps ( $Lamp_1 \sim Lamp_{10}$ ). The working flow of the MJ-level PCM circuit is that the pre-ionization circuit generates high voltage pulses with a pulse width of  $120\mu s$  through the pulse-forming component to provide energy. Then, the Xenon lamp and the pre-ionization circuit are initially triggered. After pre-ionization, the main ionization capacitor is discharged, which provides a high voltage pulse for the Xenon lamp with the pulse duration of the main ionization circuit  $470 \sim 920\mu s$ . The simulated charging voltage of pre-ionization and main ionization is 12 kV and 23kV. The specifications for the MJ-level PCM circuit are shown in Table 1.

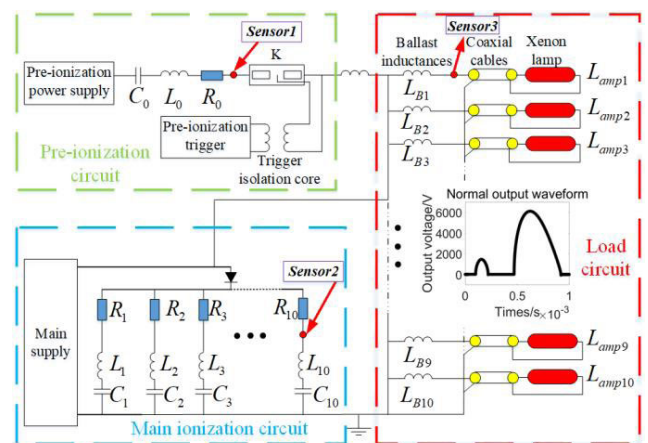


FIGURE 2. Structure of the MJ-level PCM circuit.

**TABLE 1.** Specifications of MJ-level PCM circuit simulation.

Parameters	Simulation results
The simulated charging voltage of the main ionization	23kV
Minimum delivered energy per lamp	120kJ
The pulse duration of the main ionization	470 $\mu$ s~920 $\mu$ s
The current pulse width of the main ionization	450 $\times$ (1 $\pm$ 10%) $\mu$ s
Maximum energy storage	1.2MJ
Transmission efficiency	$\geq$ 80%
The simulated voltage of pre-ionization	12kV
The pulse duration of pre-ionization	100 $\mu$ s~220 $\mu$ s
The current pulse width of pre-ionization	120 $\times$ (1 $\pm$ 10%) $\mu$ s

### 1) PRE-IONIZATION CIRCUIT

The pre-ionization circuit is composed of a pre-ionization capacitor, a pre-ionization inductor, and a pre-ionization resistor in series, with the nominal value of 15 $\mu$ F, 100 $\mu$ H, and 100m $\Omega$ , respectively. If there is no pre-ionization, the ultra-high pulse voltage of the main ionization may lead to the breakdown of the circuit components, thus affecting the stability of the MJ-level PCM circuit. The pre-ionization circuit equipped with one sensor (*sensor1*) to obtain voltage signals, and the pre-ionization pulse charges the Xenon lamps 250 $\mu$ s ahead of the main ionization pulse, forming a stable discharge channel, thus improving the efficiency of the main ionization circuit.

### 2) MAIN IONIZATION CIRCUIT AND LOAD CIRCUIT

The main ionization circuit and load circuit are equipped with *sensors2* and *sensor3*, respectively. To form the critical damping current waveform, the parameters of the main ionization circuit are relatively determined. Each branch is composed of a main ionization capacitor, a main ionization inductor, and a main ionization resistor in series, with the nominal value of 90 $\mu$ F, 150 $\mu$ H, and 200m $\Omega$ , respectively. The damping element (inductor and resistor) can not only limit the peak value of fault current but also absorb energy, which can effectively protect the main ionization circuit. The main ionization circuit has 10 groups of ballast inductors in parallel so that the current in each Xenon lamp of the load circuit are distributed equally.

## IV. PROPOSED FRAMEWORK & THEORETICAL

As shown in Fig. 3, the proposed prognosis framework for the MJ-level PCM circuit includes five steps.

**Step1: Data acquisition.** With the MJ-level PCM circuit excited by different degradation conditions, the degradation data can be acquired by adjusting the value of a circuit component to make it deviate gradually from its nominal value, and the output voltage of three sensors are gathered from RT-LAB simulator in real-time

**Step2: Feature extraction.** The time-domain features are extracted from the obtained output voltage under different degradation cycles, and the related life cycles are recorded.

**Step3: Feature selection.** A novel TSMIC algorithm is proposed to reduce the dimension of the data and remove some unrelated features.

**Step4: HI construction.** The selected feature set is used to construct HIs by the linear regression model. Additionally, the failure threshold and variation limitation in different HI curves are determined for different circuit components.

**Step5: RUP estimation.** The training feature set and HIs are inputted into the Bi-GRU network to implement model training. In the testing stage, the output of the circuit for a real signal is encoded into a testing set by the same feature extraction and feature selection methods. Ultimately, it is inputted into the trained model to obtain RUP. The following sections describe the theoretical for the five steps involved in our proposed RUP estimation framework.

### A. FEATURE EXTRACTION

For the MJ-level PCM circuit, the probability of a single fault condition is higher than that of multiple fault conditions. Therefore, this paper takes a single fault condition as a matter of course, to prove the proposed prognosis framework. A component in the circuit is randomly selected for the simulation degradation experiment, while other circuit components changed in their tolerance range. Output voltage signals of sensors are obtained, which is a set of one-dimension run-to-failure time series data. In this paper, since the output response signal of the MJ-level PCM circuit in the time domain shows a clear trend of increasing/decreasing with the degradation cycle. Hence, time-domain signal processing methods are used to extract, transform, and analyze the measured response signal data. Thus, as mentioned in Table 2, time-domain features, namely, Mean, Skewness, Margin factor (MF), Root mean square (RMS), Kurtosis, and Impulsion index (II), are used to extract the output voltage, where  $x$  is the voltage signals,  $i$  is the sample index, and  $N$  is the number of samples.

**TABLE 2.** Time-domain features.

Feature	Formula	Feature	Formula
Mean	$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$	Root mean square	$(\frac{1}{N} \sum_{i=1}^N x_i^2)^{1/2}$
Skewness	$\frac{(1/N) \sum_{i=1}^N (x_i - \bar{x})^3}{[(1/N) \sum_{i=1}^N (x_i - \bar{x})^2]^{3/2}}$	Kurtosis	$\frac{(1/N) \sum_{i=1}^N (x_i - \bar{x})^4}{[(1/N) \sum_{i=1}^N (x_i - \bar{x})^2]^2}$
Margin factor	$\frac{x_{\max}}{[(1/N) \sum_{i=1}^N  x_i ]^2}$	Impulsion index	$\frac{x_{\max}}{(1/N) \sum_{i=1}^N  x_i }$

### B. MAXIMAL INFORMATION COEFFICIENT

MIC is utilized to select sensitive features. MIC was introduced by Reshef *et al.* in 2011 [33], which aims to discover and classify the data with either linear or other functional relationships. A higher MIC value indicates that it has more dependencies relationship between the feature data, while a lower MIC value means less dependence.

The basic principle takes advantage of mutual information (MI). In order to solve the MI, the data point  $(x, y)$  is divided into  $n$  grids in  $XY$  coordinate. Giving a finite ordered pair of datasets  $C\{(x_i, y_i), i = 1, 2, \dots, n\}$ , the  $X$  and  $Y$  coordinates

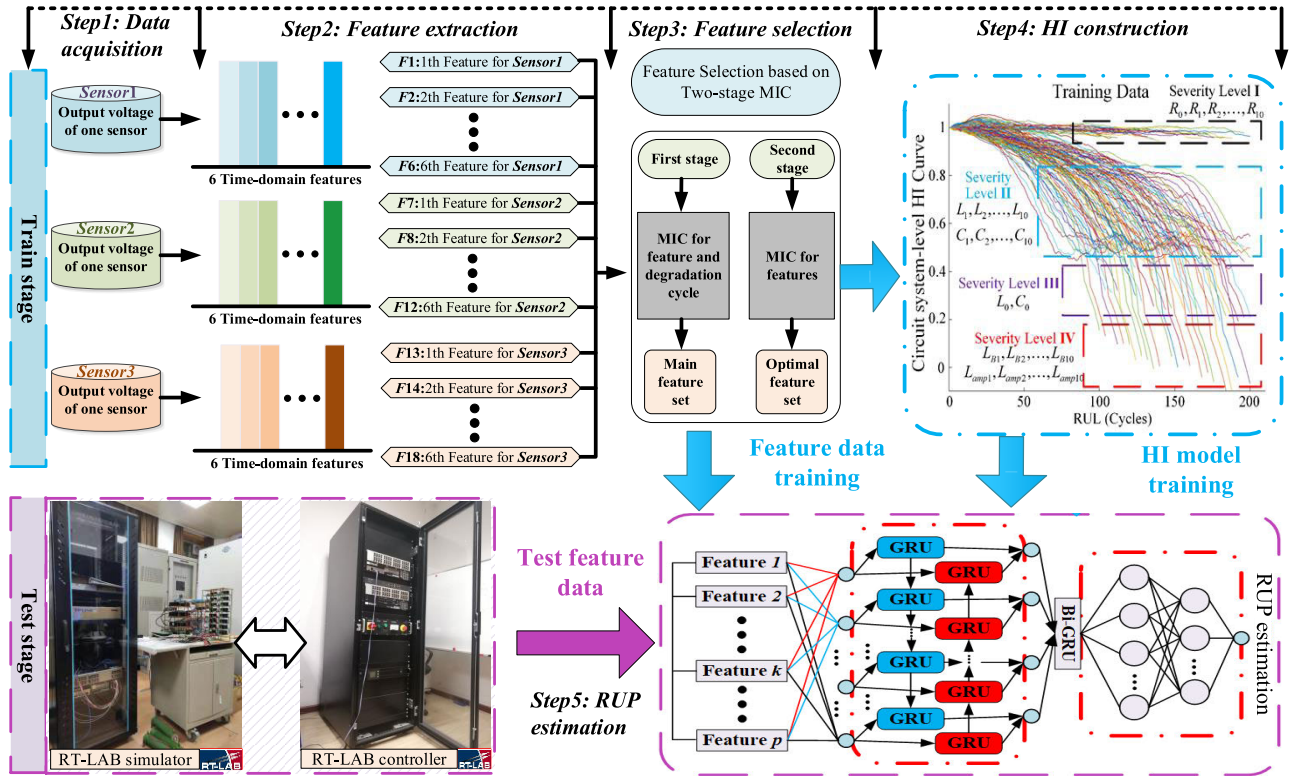


FIGURE 3. The framework of the proposed RUP estimation method.

are divided into  $s$ -by- $t$  grids, respectively. Then, grid  $T$  with shape  $s \times t$  is obtained. In practice, the number of grids is fixed, and the MI can be represented as:

$$\begin{aligned}
 I(x, y) &= H(x) + H(y) - H(x, y) \\
 &= \sum_{i=1}^s p(x_i) \log_2 \frac{1}{p(x_i)} + \sum_{j=1}^t p(y_j) \log_2 \frac{1}{p(y_j)} \\
 &\quad - \sum_{i=1}^s \sum_{j=1}^t p(x_i, y_j) \log_2 \frac{1}{p(x_i, y_j)} \\
 &= \sum_{x_i, y_j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \tag{1}
 \end{aligned}$$

where  $p(x_i, y_j)$  is the joint probability density,  $p(x_i)$  and  $p(y_j)$  are the marginal density.

In order to facilitate comparison and weighting of the values of MI on the same scale, normalization is applied to transform MI from 0 to 1, namely  $I^* \in [0, 1]$ . The element in  $s$ th row and  $t$ th column of the feature matrix  $M(C)$  on dataset  $C$  is shown below:

$$M(C)_{s,t} = \frac{I^*(C, S, T)}{\log_{\min}\{s, t\}} \tag{2}$$

Selecting the maximum value of MI in  $C$  as the value of MIC. Higher MIC values represent stronger dependency between variables, whereas lower MIC values mean weaker

correlation. The formula for MIC is showed as follows:

$$\begin{cases} MIC(C) = \max_{s*t < B(n)} \{M(C)_{s,t}\} \\ B(n) = n^{0.6} \end{cases} \tag{3}$$

where  $MIC(C)$  is the value of MIC of the two variables ( $x$  and  $y$ ) in dataset  $C$ .  $B(n)$  is a function of the sample size  $n$ , which is usually expressed as  $B(n) = n^{0.6}$  [33].

### C. HEALTH INDICATOR CONSTRUCTION

HI model reflects the evolution of the degradation in any of the circuit's critical components relevantly by utilizing multi-feature fusion according to the influence of different circuit components degradation. In this paper, multi-features are fused to form a one-dimensional vector, which is defined as the HI curve that helps in indicating the degradation trend of the circuit component. The training data  $\mathbf{Y} = [Y^{(1)}, Y^{(2)}, \dots, Y^{(u)}, \dots, Y^{(n)}]$  contains  $n$  HI curves.  $Y^{(u)} = [y_1^{(u)}, y_2^{(u)}, \dots, y_t^{(u)}, \dots, y_{L^{(u)}}^{(u)}]$  is the HI curve for simulation experiment  $u$ , where each point  $y_j^{(u)} \in \mathbf{R}^p$ . Therefore, our goal is to construct a mapping  $f: [x_p^{(1)}, x_p^{(2)}, \dots, x_p^{(u)}] \rightarrow Y^{(u)}$  between multi-feature and HI. In [46], logistic regression is proposed to fuse the multi-features into one-dimensional HIs, and then HIs are used to predict RUP through the ARMA model. However, it is found that the original degradation pattern of the MJ-level PCM circuit will be distorted by logistic regression, which may lead to inaccurate prognosis

output and larger RUP error. Consequently, this paper uses a linear regression model as a performance assessment to retain the original pattern in the selected features:

$$Y^{(u)} = \alpha + \beta^T \cdot \mathbf{x}^{(u)} + \xi = \alpha + \sum_{i=1}^p \beta_i x_i + \varepsilon \quad (4)$$

where  $\mathbf{x}^{(u)} = [x_p^{(1)}, x_p^{(2)}, \dots, x_p^{(u)}]$  is the  $p$ -dimensional selected feature vector,  $Y^{(u)}$  is the actual values of  $u$ th HI curve in the training set,  $\alpha$  is the bias,  $(\beta_1, \beta_2, \dots, \beta_p)$  is weight parameter, and  $\varepsilon$  is the noise term. Under the healthy condition, the corresponding HI value of circuit components is close to 1.

### D. BIDIRECTIONAL GATE RECURRENT UNIT NETWORK

A GRU network can only capture the dependence of the current state on the previous state (i.e., forward direction in context). In contrast, the core idea behind Bi-GRU lies in the fact that two separate hidden layers were utilized to process the sequence of feature data in two directions (i.e., forward and backward) to capture both past and future information, respectively. The Bi-GRU network is shown in Fig. 4. The following equations describe the separate hidden layers function at degradation cycle  $t$ , and two different arrow symbols,

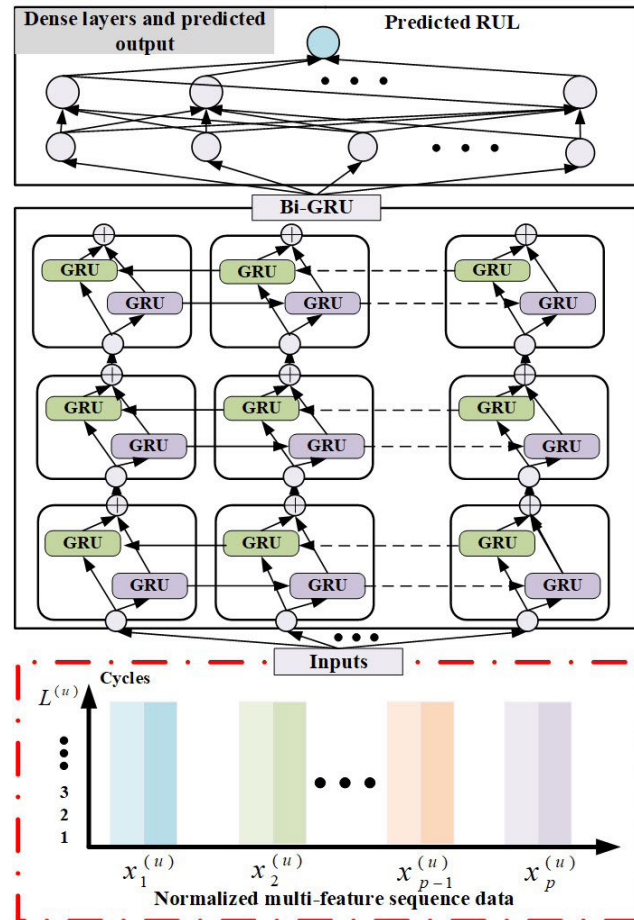


FIGURE 4. Proposed RUP estimation model.

i.e.,  $\rightarrow$  and  $\leftarrow$ , denote the forward and backward process, respectively.

$$\begin{aligned} \vec{h}_t &= f(x_t^{(u)}, \vec{h}_{t-1}; \vec{\Theta}_{Bi-GRU}) \\ &= \begin{cases} \vec{z}_t = \sigma(\vec{W}_{xz}x_t^{(u)} + \vec{W}_{hz}\vec{h}_{t-1} + \vec{b}_z) \\ \vec{r}_t = \sigma(\vec{W}_{xr}x_t^{(u)} + \vec{W}_{hr}\vec{h}_{t-1} + \vec{b}_r) \\ \vec{h}_t = \tanh(\vec{W}_{xh}x_t^{(u)}) + \vec{U}(\vec{r}_t \odot \vec{h}_{t-1}) \\ \vec{h}_t = (1 - \vec{z}_t) \odot \vec{h}_{t-1} + \vec{z}_t \odot \vec{h}_t \end{cases} \quad (5) \end{aligned}$$

$$\begin{aligned} \leftarrow{h}_t &= f(x_t^{(u)}, \leftarrow{h}_{t-1}; \leftarrow{\Theta}_{Bi-GRU}) \\ &= \begin{cases} \leftarrow{z}_t = \sigma(\leftarrow{W}_{xz}x_t^{(u)} + \leftarrow{W}_{hz}\leftarrow{h}_{t-1} + \leftarrow{b}_z) \\ \leftarrow{r}_t = \sigma(\leftarrow{W}_{xr}x_t^{(u)} + \leftarrow{W}_{hr}\leftarrow{h}_{t-1} + \leftarrow{b}_r) \\ \leftarrow{h}_t = \tanh(\leftarrow{W}_{xh}x_t^{(u)}) + \leftarrow{U}(\leftarrow{r}_t \odot \leftarrow{h}_{t-1}) \\ \leftarrow{h}_t = (1 - \leftarrow{z}_t) \odot \leftarrow{h}_{t-1} + \leftarrow{z}_t \odot \leftarrow{h}_t \end{cases} \quad (6) \end{aligned}$$

where  $\vec{\Theta}_{Bi-GRU}$  and  $\leftarrow{\Theta}_{Bi-GRU}$  are the parameters set of the forward and backward process, which are shared by all degradation cycles and learned during model training.  $h_{t-1}$  represents the hidden states of the degradation cycle ( $t-1$ ).  $\vec{b}_z, \vec{b}_r$  and  $\leftarrow{b}_z, \leftarrow{b}_r$  are bias weights of the forward and backward process.  $\vec{W}_{xz}, \vec{W}_{hz}, \vec{W}_{xr}, \vec{W}_{hr}, \vec{W}_{xh}, \vec{U}$  and  $\leftarrow{W}_{xz}, \leftarrow{W}_{hz}, \leftarrow{W}_{xr}, \leftarrow{W}_{hr}, \leftarrow{W}_{xh}, \leftarrow{U}$  are input weights of the forward and backward process.  $\sigma$  and  $\tanh$  (hyperbolic tangent) are pointwise nonlinear activation functions and  $\odot$  denote pointwise multiplication of two vectors.

In the first stage of RUP estimation, the input of the Bi-GRU network includes selected features and HIs, which compose the training set  $\{\mathbf{x}^{(u)}, Y^{(u)}\}_{u=1}^n$ . Therein, the shape of the input feature  $\mathbf{x}^{(u)} = [x_p^{(1)}, x_p^{(2)}, \dots, x_p^{(u)}]$  is  $L^{(u)} \times p$ .  $p$  represents the number of sensitive features selected by the TSMIC algorithm, and  $L^{(u)}$  denotes the length of the last degradation cycle at  $u$ th simulation experiment. The input  $Y^{(u)} = [y_1^{(u)}, y_2^{(u)}, \dots, y_t^{(u)}, \dots, y_{L^{(u)}}^{(u)}]$  indicates the actual value of HIs in the training set. Then, the output functions are calculated as follows:

$$\begin{aligned} H_u &= [h_1, h_2, \dots, h_t, \dots, h_p] \\ &= f(\mathbf{x}^{(u)}; \vec{\Theta}_{Bi-GRU}, \leftarrow{\Theta}_{Bi-GRU}) \quad (7) \end{aligned}$$

where  $f$  denotes the hidden layer function of the Bi-GRU network, which is defined by Eq. (5) and (6).  $H_u$  represents the  $u$ th sequence of the output functions processed by the first stage of the Bi-GRU network, which is characterized by the parameters set  $(\vec{\Theta}_{Bi-GRU}$  and  $\leftarrow{\Theta}_{Bi-GRU})$ . The complete output  $h_t$  at the degradation cycle  $t$  is the sum of the output elements from the forward and backward process, which can be calculated:

$$h_t = \vec{h}_t \oplus \leftarrow{h}_t \quad (8)$$

where  $\oplus$  represents the sum of the elements of the two vectors.

In the second stage of RUP estimation, the output from the first stage is inputted to the fully connected layers (FCs) to

seek a higher level of representation, which can be defined by the following equation:

$$o_u = f(H_u; \Theta_{FC}) = g(\mathbf{W}_u h_p + \mathbf{b}_u) \quad (9)$$

where  $o_u$  is the output vector of a FC.  $\Theta_{FC}$  represents the parameters set, i.e., weight matrix  $\mathbf{W}_u$  and bias vector  $\mathbf{b}_u$ .  $g(\cdot)$  denotes the activation function of the neurons, which is set as a rectified linear unit. Dropout is adopted in the Bi-GRU network to solve the problem of overfitting. A rectified linear unit (ReLU) is used in the network. Ultimately, the final learned output features of the FC layers are put into the linear regression layer, and the prognosis results are generated.

$$\hat{y}_u = \mathbf{W}_o o_u \quad (10)$$

where  $\hat{y}_u$  is the predicted RUP, and  $\mathbf{W}_o$  denotes the weight vector of the final linear regression layer.

## V. IMPLEMENTATION RESULTS AND DISCUSSION

### A. EXPERIMENTAL SETUP

In this paper, numerous cases are conducted with MJ-level PCM circuit simulation data (provided by CAEP, simulated in RT-LAB hardware-in-the-loop simulator) to verify the effectiveness of the proposed method. RT-LAB hardware-in-the-loop simulator is an industrial real-time simulation platform developed by Opal-RT technologies in Canada. The MATLAB/Simulink model of the tested circuit is applied to the RT-LAB hardware-in-the-loop simulation system by PC, which reduces the difficulty of constructing the circuit and improves the reliability of the simulation system. Additionally, the data processing methods mentioned are implemented with MATLAB R2019a. The RUP estimation methods are a server with Intel Xeon CPU E5-2698 v4 2.2-GHz, 64-GB RAM. The programming language is Python 3.5 with a deep learning library “Keras” (2.0.8) using open-source software library “TensorFlow” (1.3.0) as backend.

As illustrated by Fig. 5, the simulation experimental of the MJ-level PCM circuit was built in the OP5600 simulator, which constructs a circuit response database containing multiple degradation conditions and transmits the multi-sensors voltage signal to the PC. The circuit response was captured at the output using a National Instruments (NI) USB-6212 data acquisition board. The data were recorded using LabVIEW on PC. The experiment operations and degradation parameter setting of different circuit components are implemented in the OP5607 controller. After collecting the circuit degradation data, the proposed data-driven approach based TSMIC and Bi-GRU algorithm are devised for tracking the evolution of the HI and predicting the MJ-level PCM circuit’s RUP.

### B. DEGRADATION DATA ACQUISITION

**Tolerance range, Failure Threshold, and Variation Limitation:** The relationship among tolerance range, failure threshold, and variation limitation should be predefined. There are two mainstream viewpoints in the existing literature: 1) One viewpoint is that circuit components with  $\pm 30\%$  deviation

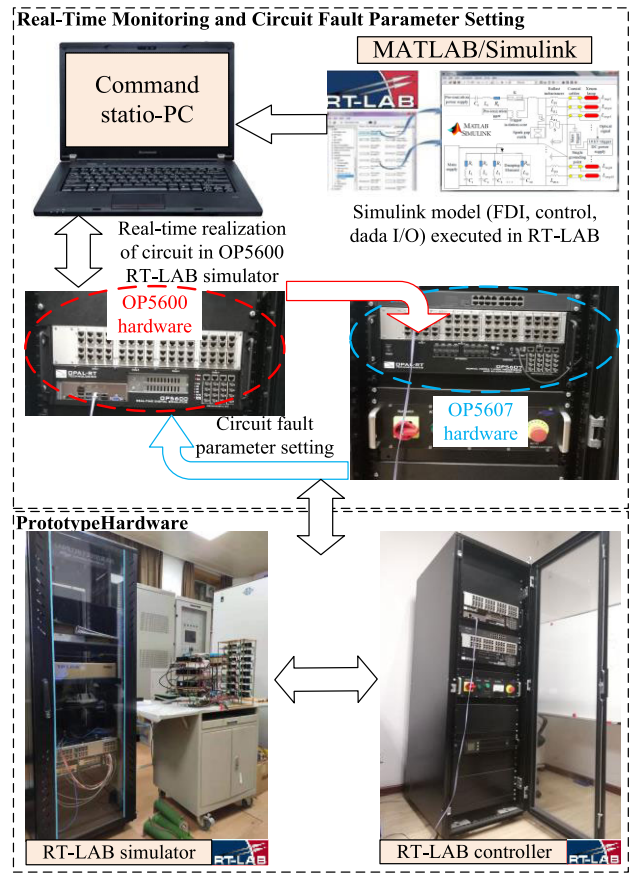


FIGURE 5. Real-time realization of the MJ-level PCM circuit with RT-LAB hardware-in-the-loop simulation.

from its nominal value (**regardless of its tolerance range**) should be regarded as failure [14], [47]–[49]. 2) Another viewpoint is that when the components of a circuit deviate beyond its tolerance range, the circuit is regarded as a failure. According to the characteristics of different circuit components, it is divided into the following three cases: a tolerance of  $\pm 1\%$  for all the components [50],  $\pm 5\%$  for resistors, and  $\pm 10\%$  for capacitors adopted in previous work [17], or  $\pm 10\%$  for both resistors and capacitors. Due to a large number of components in the MJ-level PCM circuit, including resistors, capacitors, inductors, and xenon lamps. It is necessary to consider both the failure threshold and tolerance range.

Therefore, this paper adopts the viewpoint that the MJ-level PCM circuit reaches failure if the component value has exceeded a predefined failure threshold ( $\pm 30\%$  deviation from its nominal value). Since the value of circuit components cannot be increased or decreased unlimited, it is considered that the variation limitation of each circuit component is  $\pm 40\%$  deviates from its nominal value. Furthermore, in this paper, according to the influence of degradation for different circuit components on the output voltage waveforms distortion, which is shown in the *Section IV-A-(2)*. The tolerance range of components are divided into four cases according to the importance of components of the MJ-level PCM circuit:  $\pm 10\%$  for resistors,  $\pm 5\%$  for main



ionization capacitors/inductors,  $\pm 3\%$  for pre-ionization capacitors/inductors, and  $\pm 1\%$  for Xenon lamps/ballast inductors, respectively.

### 1) DEGRADATION PARAMETER SETTING

According to the above definition, the variation limitation and failure threshold of the circuit components are obtained.

$$\begin{cases} \text{Value}_{\text{Limitation\_max}} = 1.4 \times \text{Value}_0, & \text{if } \uparrow \\ \text{Value}_{\text{Limitation\_min}} = 0.6 \times \text{Value}_0, & \text{if } \downarrow \\ \text{Value}_{\text{Failure\_max}} = 1.3 \times \text{Value}_0, & \text{if } \uparrow \\ \text{Value}_{\text{Failure\_min}} = 0.7 \times \text{Value}_0, & \text{if } \downarrow \end{cases} \quad (11)$$

where  $\uparrow$  and  $\downarrow$  indicate the increase and decrease of the value of the circuit component.  $\text{Value}_0$  is the nominal value of the circuit component.  $\text{Value}_{\text{Limitation}}$  and  $\text{Value}_{\text{Failure}}$  are the value of variation limitation and failure threshold of the component. The number of degradation cycle for failure threshold and variation limitation is defined as follows

$$\begin{cases} \text{Cycle}_{\text{Failure}} = |\text{Value}_{\text{Failure}} - \text{Value}_0| / \text{Value}_1 \\ \text{Cycle}_{\text{Limitation}} = |\text{Value}_{\text{Limitation}} - \text{Value}_0| / \text{Value}_1 \end{cases} \quad (12)$$

where  $\text{Value}_1$  represents the variation step of the experimental component with each degradation cycle. The value of each component increases/decreases equally and gradually concerning the degradation cycle.

The degradation parameters of the MJ-level PCM circuit are recorded in Table 3. Take  $R_1$  as an example, the nominal value of  $R_1$  is 200k $\Omega$ . Since the variation limitation of each circuit component is  $\pm 40\%$  deviates from its nominal value, the upper value of variation limitation and lower value of variation limitation are 280k $\Omega$  and 120k $\Omega$ , respectively. If the value of  $R_1$  increases by 0.8k $\Omega$  in each variation step, the number of degradation cycles is 100 from 200 k $\Omega$  to 280 k $\Omega$ . If the value of  $R_1$  increases by 0.4k $\Omega$  in each variation step, the number of degradation cycles is 200 from 200 k $\Omega$  to 280 k $\Omega$ . Besides, the upper value of failure threshold/ tolerance range and lower value of failure threshold/tolerance range are 260k $\Omega$ /220k $\Omega$  and 140k $\Omega$ /180k $\Omega$ , respectively. The other components are operated in the same way.

**TABLE 3.** The parameters of MJ-level PCM circuit simulation.

Failure Class	Nominal Value	Tolerance Range	Failure Threshold	Variation Limitation	Variation Step
$(R_{1-10})\uparrow$	200k $\Omega$	220k $\Omega$	260k $\Omega$	280k $\Omega$	(0.4–0.8)k $\Omega$
$(R_{1-10})\downarrow$	200k $\Omega$	180k $\Omega$	140k $\Omega$	120k $\Omega$	(0.4–0.8)k $\Omega$
$(C_{1-10})\uparrow$	90 $\mu$ F	94.5 $\mu$ F	117 $\mu$ F	126 $\mu$ F	(180–360)nF
$(C_{1-10})\downarrow$	90 $\mu$ F	85.5 $\mu$ F	63 $\mu$ F	54 $\mu$ F	(180–360)nF
$(L_{1-10})\uparrow$	150 $\mu$ H	157.5 $\mu$ H	195 $\mu$ H	210 $\mu$ H	(300–600)nH
$(L_{1-10})\downarrow$	150 $\mu$ H	142.5 $\mu$ H	105 $\mu$ H	90 $\mu$ H	(300–600)nH
$R_0\uparrow$	100k $\Omega$	110k $\Omega$	130k $\Omega$	140k $\Omega$	(0.2–0.4)k $\Omega$
$R_0\downarrow$	100k $\Omega$	90k $\Omega$	70k $\Omega$	60k $\Omega$	(0.2–0.4)k $\Omega$
$C_0\uparrow$	15 $\mu$ F	15.45 $\mu$ F	19.5 $\mu$ F	21 $\mu$ F	(30–60)nF
$C_0\downarrow$	15 $\mu$ F	14.55 $\mu$ F	10.5 $\mu$ F	9 $\mu$ F	(30–60)nF
$L_0\uparrow$	100 $\mu$ H	103 $\mu$ H	130 $\mu$ H	140 $\mu$ H	(0.2–0.4) $\mu$ H
$L_0\downarrow$	100 $\mu$ H	97 $\mu$ H	70 $\mu$ H	60 $\mu$ H	(0.2–0.4) $\mu$ H
$\text{Lamp}_{(1-10)}\uparrow$	K=100	101	130	K=140	0.2–0.4
$\text{Lamp}_{(1-10)}\downarrow$	K=100	99	70	K=60	0.2–0.4
$L_{B(1-10)}\uparrow$	30 $\mu$ H	30.3 $\mu$ H	39 $\mu$ H	42 $\mu$ H	(30–60)nH
$L_{B(1-10)}\downarrow$	30 $\mu$ H	29.7 $\mu$ H	21 $\mu$ H	18 $\mu$ H	(30–60)nH

### 2) MULTI-SENSORS OUTPUT VOLTAGE

For the MJ-level PCM circuit, the degradation of different components has different effects on the output voltage.

#### a: MAIN IONIZATION CAPACITORS/INDUCTORS

The failure degradation of main ionization capacitors ( $C_1 \sim C_{10}$ ) and inductors ( $L_1 \sim L_{10}$ ) only affects the circuit output voltage waveform of main ionization, and the main ionization circuit has ten branches, one component failure has less influence on the circuit system. Fig. 6. (a) and (b) show the output voltage response corresponding to different values of  $C_1$  and  $L_1$ , where 0,  $-20\%$  and  $20\%$  represent the output voltage at a nominal value, a  $20\%$  increase in nominal value and a  $20\%$  decrease in nominal value, respectively.

#### b: PRE-IONIZATION CAPACITORS/INDUCTORS

The Pre-ionization circuit plays an essential role in transition ionization. Moreover, the pre-ionization circuit has only one branch. The degradation of the pre-ionization capacitor and inductor will affect the output voltage of the pre-ionization circuit. Fig. 6 (c) and (d) show the output voltage response corresponding to different values of  $C_0$  and  $L_0$ .

#### c: XENON LAMPS AND BALLAST INDUCTORS

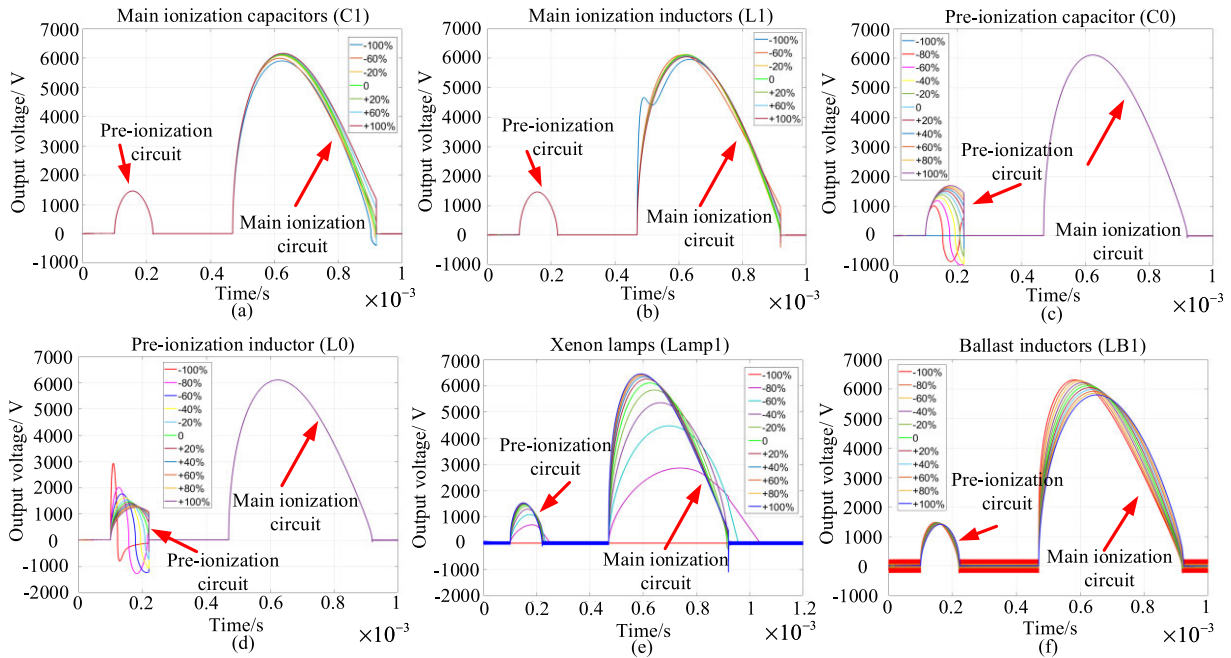
The arc plasma produced by the discharge of the Xenon lamp, which has a strong dependence on the load current, and it represents a nonlinear dynamic resistor characteristic. The resistor coefficient of Xenon lamp is  $K = 1.27(p/450)^{0.2}(4l/d)$ , where  $p$  is the internal pressure of the Xenon lamp,  $l$  is the arc length of the Xenon lamp,  $d$  is the inner diameter of the lamp tube. According to the Gonze model, the nonlinear dynamic resistor of Xenon lamp can be expressed as

$$R = K i^{(-1/2)} \quad (13)$$

The resistor coefficient of the Xenon lamp has been set as  $K = 100$  in this paper. The degradation of the Xenon lamp is simulated by adjusting the value of  $K$ . Furthermore, the function of the ballast inductor is to distribute the current evenly in each branch of the load circuit. Fig. 6 (e) and (f) show the output voltage response corresponding to the  $K$  value deviation of the Xenon lamp $_1$  and different values of ballast inductor  $L_{B1}$ . Once the performance of Xenon lamps ( $\text{Lamp}_1 \sim \text{Lamp}_{10}$ ) and ballast inductors ( $L_{B1} \sim L_{B10}$ ) are degraded, the main ionization and pre-ionization voltage waveforms will be distorted.

### C. EXTRACTING TREND FEATURES

To facilitate the statistics of degradation rules of each circuit component, the number of degradation cycles in each simulation experiment ranges from 100 to 200. What's more, there are 53 components in the MJ-level PCM circuit. In order to cover the degradation data of each component as much as possible, 3 and 2 degradation experiments are set for each selected component in the training set and testing set, respectively. Consequently, 159 times and 106 times simulation experiments are carried out on different components



**FIGURE 6.** Output voltage response to value variation of the component. (a) Main ionization capacitor  $C_1$ ; (b) Main ionization inductor  $L_1$ ; (c) Pre-ionization capacitor  $C_0$ ; (d) Pre-ionization inductor  $L_0$ ; (e)  $K$  value of Xenon Lamp $_1$ ; (f) Ballast inductor  $LB_1$ .

**TABLE 4.** The basic information of training data and testing data.

Dataset	Simulation experiment	Max. cycle	Min. cycle	Fault components	No. feature
Training data	159 times	200	100	53	18
Testing data	106 times	200	100	53	18

in different cycles for the training set and testing set. The essential information is found in Table 4.

There are 18 features in total, which are further divided into three sensors condition (*sensor1*, *sensor2*, *sensor3*), and their details are shown in Table 5. For example, *S1\_RMS* represents the root mean square of *sensor1*'s output voltage. More formally, as shown in Fig. 7, for each simulation experiment  $u \in [1, 2, \dots, 159]$ , the  $m$ th ( $m \in [1, 2, \dots, 18]$ ) feature is an one-dimensional time-series data, which represented as  $x_t^m = [C_1^m, C_2^m, \dots, C_t^m, \dots, C_{L(u)}^m]$ , where the  $L(u) \in [100, 200]$  represents the length of the last degradation cycle. Each point  $C_t^m \in \mathbf{R}^m$  is a time-domain feature for output voltage at degradation cycle  $t$ . For 159 times simulation experiments, the  $m$ th feature set is a 159-dimensional multivariate time-series data, which is represented as  $F_m = [x_m^{(1)}, x_m^{(2)}, \dots, x_m^{(159)}]$ . The raw training feature is recorded as  $F = [F_1, F_1, \dots, F_m, \dots, F_{18}]$ .

Fig. 8. shows the partially extracted features in the training set, and it can be observed that different features have different responses to components degradation. Every feature contains 159 multivariate time series data, and the cycle number ranges from 100 to 200. However, not all the features well indicate the variation of health degradation, which indicates that these features are invalid and may not perform well in

**TABLE 5.** Time-domain features and feature codes.

Code	Feature	Code	Feature	Code	Feature
F1	<i>S1_Mean</i>	F7	<i>S2_Mean</i>	F13	<i>S3_Mean</i>
F2	<i>S1_Skewness</i>	F8	<i>S2_Skewness</i>	F14	<i>S3_Skewness</i>
F3	<i>S1_MF</i>	F9	<i>S2_MF</i>	F15	<i>S3_MF</i>
F4	<i>S1_RMS</i>	F10	<i>S2_RMS</i>	F16	<i>S3_RMS</i>
F5	<i>S1_Kurtosis</i>	F11	<i>S2_Kurtosis</i>	F17	<i>S3_Kurtosis</i>
F6	<i>S1_II</i>	F12	<i>S2_II</i>	F18	<i>S3_II</i>

RUP estimation. As shown in Fig. 8. (a) and (b), these features are poorly correlated with the degradation cycle, which is removed in the first stage of feature selection. The features in Fig. 8. (c) and (d) present different attenuation curves with a lot of clutters. These features are poorly correlated with other features, which are deleted in the second stage of feature selection. Reasonable features have a good correlation with the process of health degradation, which shows a trend of monotonous increase or decrease. Therefore, the features Fig. 8. (e) and (f) represent the regular degradation trend, which will improve the prognosis accuracy.

#### D. FEATURE SELECTION BASED ON TSMIC

In this part, the TSMIC algorithm is proposed to overcome the shortcoming of insufficient consideration of nonlinear relations and mine the deep mutual information between features. In the first stage, the main feature subset, where features hardly change with the degradation cycle are eliminated, can be obtained by calculating the values of MIC between every feature and the degradation cycle. If the feature does not satisfy Eq. (14), it will be removed.

$$MIC_i(F_i, t) \geq \sigma_1 \tag{14}$$

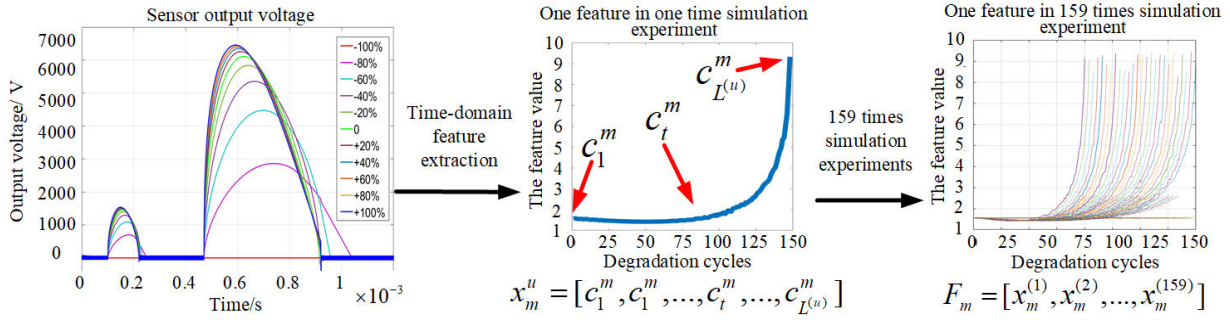


FIGURE 7. The process of extracting trend features.

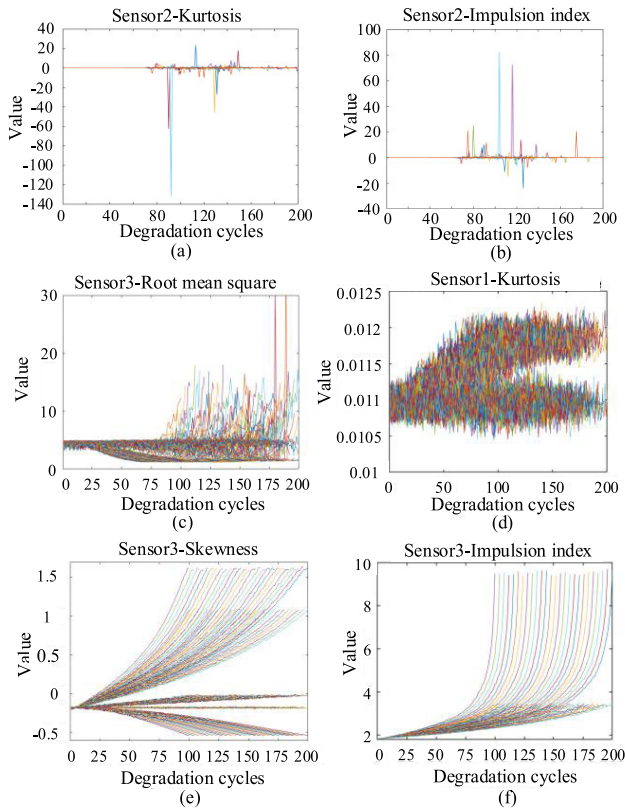


FIGURE 8. Extracted features for training data. (a) *Sensor2\_Kurtosis*; (b) *Sensor2\_Impulsion index*; (c) *Sensor1\_RMS*; (d) *Sensor1\_Kurtosis*; (e) *Sensor3\_Skewness*; (f) *Sensor3\_Impulsion index*.

where  $F_i$  denotes the  $i$ th feature,  $t$  is the cycle of degradation,  $MIC_i$  represents the value of MIC between  $i$ th feature and degradation cycle  $t$ .  $\sigma_1$  is the threshold in the first stage, which is expressed as:

$$\sigma_1 = \frac{1}{N} \sum_{i=1}^N MIC_i(F_i, t), \quad i = 1, 2, \dots, N \quad (15)$$

where  $N$  represents the number of features in the raw feature set  $F = [F_1, F_2, \dots, F_{18}]$ . As shown in Fig. 9, using the feature labels as a horizontal coordinate and the corresponding MIC values of the features as the ordinate and the threshold

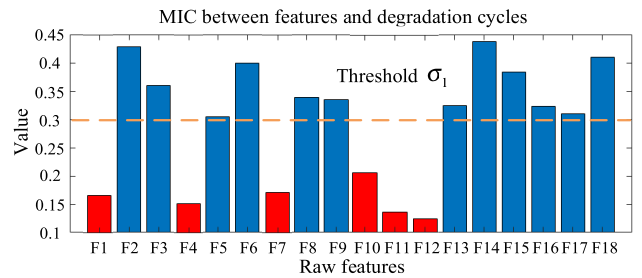


FIGURE 9. The value is ranking between features and degradation cycles based on MIC.

$\sigma_1$  is calculated as 0.3. Finally, for the ease of eliminating the features which hardly change with the degradation cycle, the main feature set is regarded as  $S = [F_2, F_3, F_5, F_6, F_8, F_9, F_{13}, F_{14}, F_{15}, F_{16}, F_{17}, F_{18}]$ .

In the second stage of feature selection, the values of MIC between any two features in the main feature set  $S$  are calculated. Then, a symmetric matrix is established, where  $m_{jk}$  denotes the value of MIC between the  $j$ th feature and  $k$ th feature, and all the diagonal values are 1. Due to the symmetry of the matrix, the average MIC per line, namely  $Mean = (Mean_1, \dots, Mean_j, \dots, Mean_k)$ , need be considered.  $Mean_j$  can reflect the degree of correlation between all other features with the  $j$ th feature. The matrix is as follows:

$$H = \begin{bmatrix} 1(m_{11}) & \cdots & m_{1j} & \cdots & m_{1k} & \vdots & Mean_1 \\ \vdots & \ddots & \vdots & & \vdots & \vdots & \vdots \\ m_{j1} & & 1(m_{jj}) & & m_{jk} & \vdots & Mean_j \\ \vdots & & \vdots & \ddots & \vdots & \vdots & \vdots \\ m_{k1} & \cdots & m_{kj} & \cdots & 1(m_{kk}) & \vdots & Mean_k \end{bmatrix} \quad (16)$$

The symmetric matrix in Eq. (16) is shown in Fig. 10, where each value represents the value of MIC between two features in the main feature set  $S$ . The higher the value of correlation, the better the feature representing the declining trend of circuit health status. The average MIC  $Mean_i$  in per line should be considered to eliminate the redundant

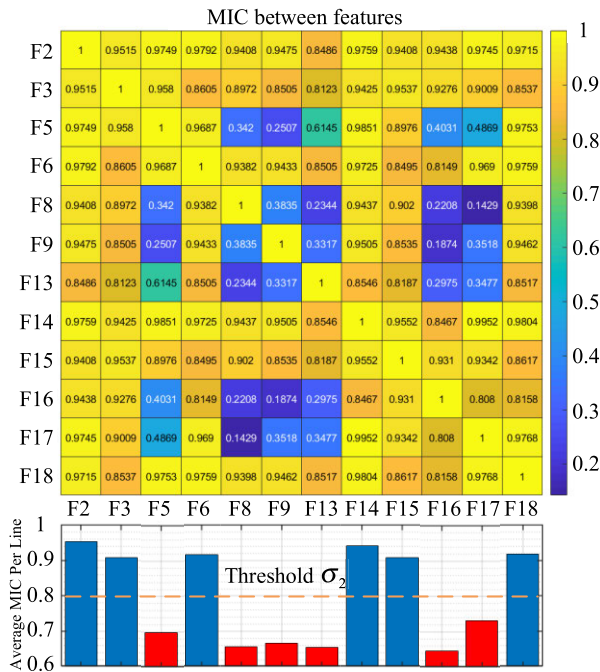


FIGURE 10. The values of MIC between features of the main feature set.

features, which can reflect the degree of correlation between all other features with the  $i$ th feature. Then,  $Mean_j$  satisfies the Eq. (17), the  $j$ th feature is regarded as the optimal feature. Otherwise, it will be eliminated.

$$Mean_j \geq \sigma_2 \quad (17)$$

where  $Mean_j$  represents the average MIC of  $j$ th row.  $\sigma_2$  is the threshold in the second stage, which is expressed as:

$$\sigma_2 = \frac{1}{M} \sum_{i=1}^M Mean_j \quad (18)$$

where  $M$  is the number of features in the main feature set  $S$ .

For the off-diagonal elements, if  $m_{ij} \geq \sigma_2$ , ( $i \leq j$ ), delete the feature (set  $m_{ij}$ ,  $m_{jj} = 0$ ), the matrix is updated as follows:

$$H' = \begin{bmatrix} 1(m_{11}) & \cdots & 0 & \cdots & m_{1k} \\ & \ddots & \vdots & & \vdots \\ & & 0 & & m_{jk} \\ & & & \ddots & \vdots \\ & & & & 1(m_{kk}) \end{bmatrix} \quad (19)$$

From Eq. (18), the threshold  $\sigma_2$  is calculated as 0.8. If  $m_{ij}$  does not satisfy Eq. (18), the corresponding diagonal value  $m_{jj}$  becomes 0. Subsequently, all diagonal features with the value of 1 are selected as the optimal features with the strongest correlation. Finally, after feature selection through a TSMIC algorithm, the optimal features are represented as  $P = [F_2, F_3, F_6, F_{14}, F_{15}, F_{18}]$ .

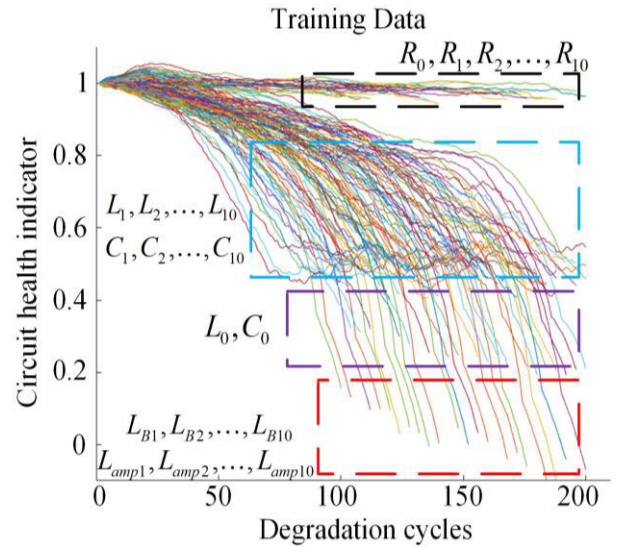


FIGURE 11. Trend exhibited by the HI during the degradation of circuit components.

### E. HEALTH INDICATOR CURVES

The HI curve is set relevant to each circuit component, then followed by tracking its HI value to predict the end of performance (EOP) of the components in the MJ-level PCM circuit. In this paper, HI curves are calculated by Eq. (4) and fitted with the linear regression model of health condition. Then, a training HI model  $\{M_{(u)}\}$  can be established to describe the degradation performance of circuit components from normal to failure:

$$M_{(u)} : y_t^{(u)} = f_{(u)}(t), \quad 0 \leq t \leq L^{(u)}, \quad u = 1, 2, \dots, 159 \quad (20)$$

where  $y_t^{(u)}$  is a point data at the degradation cycle  $t$  for one HI curve,  $L^{(u)}$  is the length of the last degradation cycle before circuit component failure for simulation experiment  $u$ . HI model  $\{M_{(u)}\}$  can be trained by the Bi-GRU network for producing an estimated output. Fig. 11 shows the 159 HI curves for different components in the MJ-level PCM circuit, which are constructed by the linear regression model. Therein, one of the HI curves indicates the degradation trend for one of the components deviated from their nominal values.

### F. FAILURE PROGNOSTICS

#### 1) PROGNOSTIC PERFORMANCE METRICS

The error in estimating the RUP of the  $n$ th circuit component is given by

$$E_n = RUP_{Estimated} - RUP_{Actual} \quad (21)$$

where  $RUP_{Estimated}$  and  $RUP_{Actual}$  denote the estimated RUP and actual RUP, respectively. And  $n$  represents the  $n$ th testing circuit component.  $E_n$  can be divided into early prediction (i.e., the estimated RUP value is smaller than the actual RUP value) and late prediction (i.e., the estimated RUP value is larger than the actual RUP value).

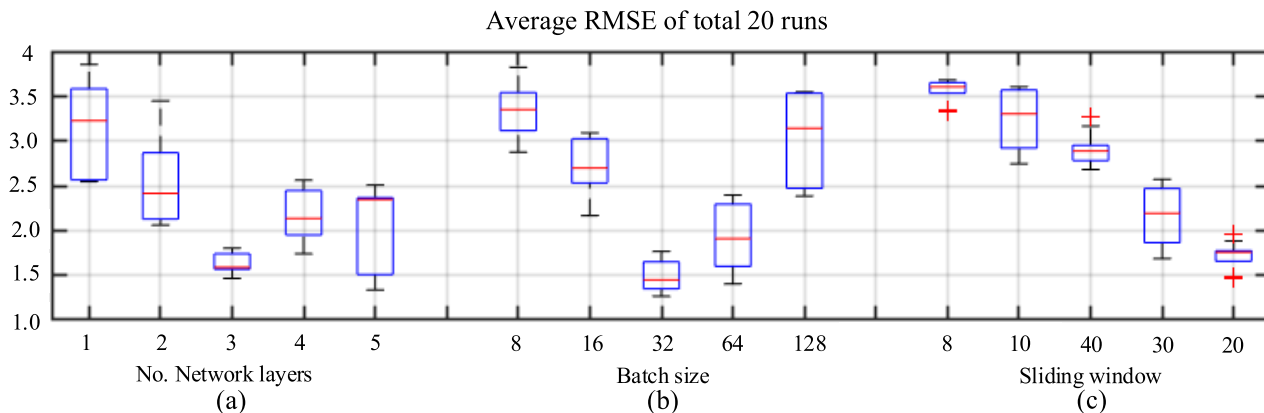


FIGURE 12. The effect of RUP estimation for different parameters in Bi-GRU (a) Number of network layers; (b) Batch size; (c) Sliding window.

In this work, the root mean square error (RMSE) is used to evaluate the prognosis accuracy of RUP. A smaller RMSE error reflects the higher effectiveness and stability of the prognosis result. The formulation of RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (E_n)^2} \quad (22)$$

## 2) THE PARAMETERS SELECTION IN BI-GRU MODEL

The network structure has a significant impact on the performance of prognosis results. Considering that the prediction results may be affected by many factors, such as parameter initialization, and the dropout technique, each result is the average of 10 repeated experiments.

### a: EFFECTS OF THE DIFFERENT NUMBERS OF THE NETWORK LAYER

The training time increase with the network structure deepening, and there will be the risk of overfitting due to the limited number of training samples in the experiment. Thus, the number of total hidden layers is constrained between 1 and 5. As shown in Fig. 13 (a), the structure with three hidden layers can produce the best results of RUP estimation among a total of 10 predictions when compared with the rest of the structures. Therefore, it is used as the default network structure in the subsequent simulation experiment.

### b: EFFECTS OF DIFFERENT BATCH SIZES

Another critical parameter in the network is the batch size. Fig. 12 (b) shows the results of RUP estimation in different batch sizes. Because of the importance of early prediction needing more consideration in condition-based maintenance, batch size with a smaller value of RMSE is selected within a certain error range. The best performance is acquired when the batch size sets to 32.

TABLE 6. List of hyper-parameters for Bi-GRU model.

Hyper-parameters	Description	Suggested value
Batch size	Number of training samples in one forward/backward pass	32
Dropout rate	Probability of dropping out units in the Bi-GRU network	0.2
No. Network layer	Number of network layers in Bi-GRU	3
Steps in the sliding window	The sliding window is used to smooth the signals	20
Training epochs	Number of epochs for forward/backward pass	100
Early stopping criteria	Number of epochs with no improvement in testing error	2

### c: EFFECTS OF THE DIFFERENT STEPS OF THE SLIDING WINDOW

The step of sliding window decides how much information to be trained in one sample preparation. This work presents the effect of the sliding window size in Fig. 12 (c). When the step of the sliding window is set as 20, RMSE results are the lowest. On the one hand, the Bi-GRU network cannot extract effective features hidden in shorter sequence degradation data, which contains limited sequence information. On the other hand, although more extended sequence degeneration data contain more local information, longer steps can increase the overfitting risk of the Bi-GRU network during the training stage.

For the choice of the hyperparameters such as the number of the network layer, batch size, and step of sliding window dropout rate, training epochs, and early stopping criteria, we implemented a grid search strategy to obtain a satisfactory prediction performance. The hyperparameters in the Bi-GRU are summarized in Table 6.

## 3) PROGNOSIS RESULTS OF THE PROPOSED METHOD

When the prognosis module is triggered, the HI curve is tracked at any degradation cycle to predict the RUP of the

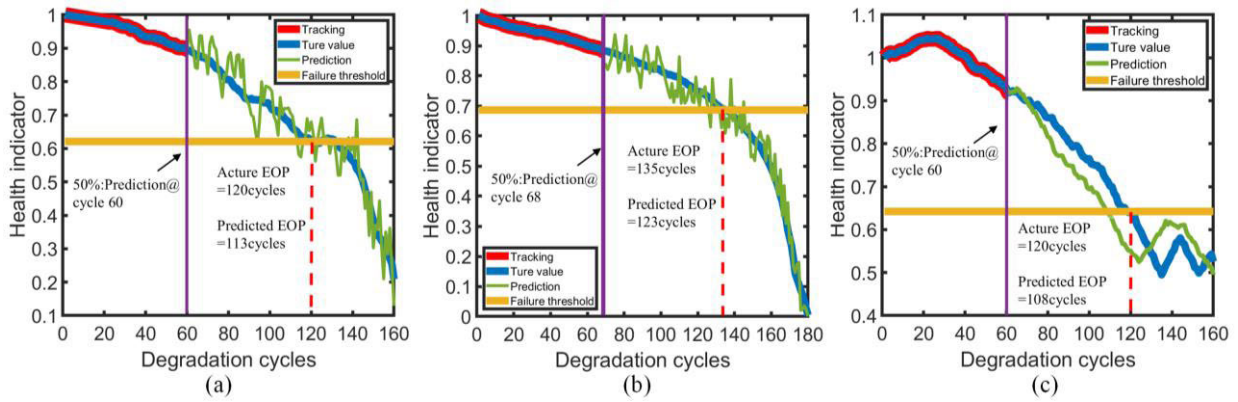


FIGURE 13. RUP estimation at degradation cycle 60, 68, 60 for (a) Pre-ionization capacitor  $C_0$ ; (b) Xenon Lamp<sub>8</sub>; (c) Main ionization inductor  $L_9$ , respectively.

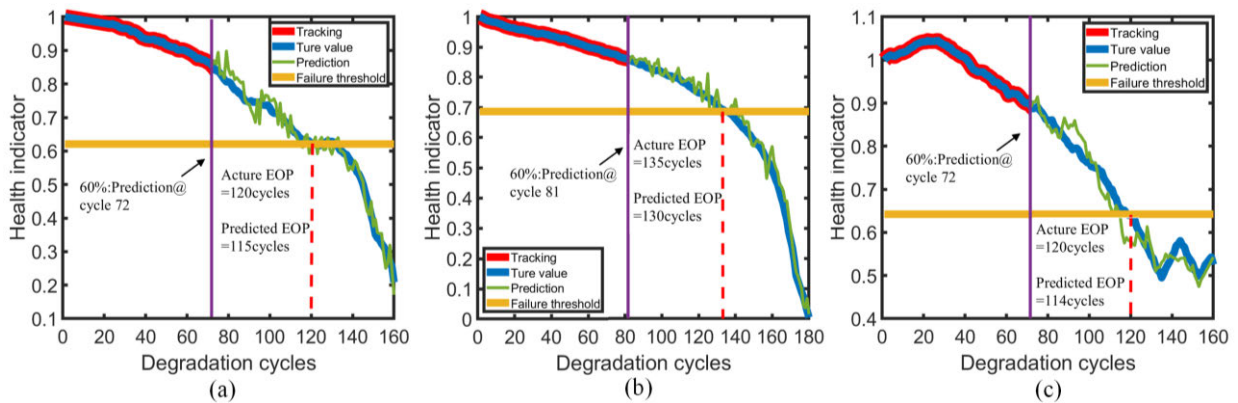


FIGURE 14. RUP estimation at degradation cycle 72, 81, 72 for (a) Pre-ionization capacitor  $C_0$ ; (b) Xenon Lamp<sub>8</sub>; (c) Main ionization inductor  $L_9$ , respectively.

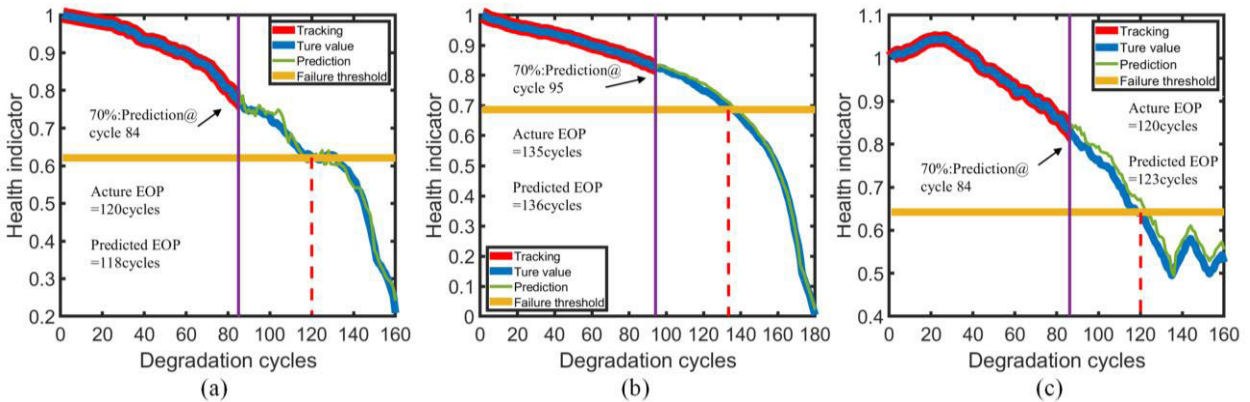


FIGURE 15. RUP estimation at degradation cycle 84, 95, 84 for (a) Pre-ionization capacitor  $C_0$ ; (b) Xenon Lamp<sub>8</sub>; (c) Main ionization inductor  $L_9$ , respectively.

circuit component. As shown in Fig. 13. (a)-(c), variation limitation in a pre-ionization capacitor  $C_0$  (test experiment #31), a Xenon Lamp<sub>8</sub> (test experiment #61), and a main ionization inductor  $L_9$  (test experiment #76) are 160, 180 and 160, which represent the value of  $C_0$ , Lamp<sub>8</sub> and  $L_9$  increase/decrease 40% from their nominal value. The degradation for each component was assumed to increase/decrease gradually and evenly concerning the degradation cycle. Thus, the failure

threshold in the  $C_0$ , Lamp<sub>8</sub>, and  $L_9$  are 120 ( $160 \times 3.4$ ), 135 and 120, which represent the value of  $C_0$ , Lamp<sub>8</sub>, and  $L_9$  increase/decrease 30% from their nominal value. In addition, the prognosis experiments for  $C_0$ , Lamp<sub>8</sub>, and  $L_9$  at degradation cycle 60 ( $120 \times 50\%$ ), 68, 60, which is truncated at 50% of the testing run to failure data. More precisely, take  $C_0$  as an example, Fig. 13(a) shows that RUP estimation for  $C_0$  starts at degradation cycle 60. It is seen that tracking and

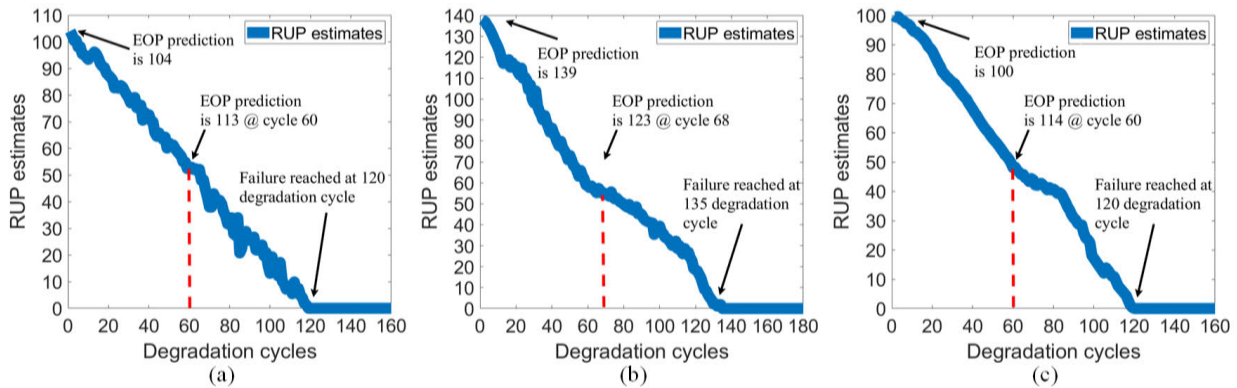


FIGURE 16. RUP estimation at every degradation cycle for (a) Pre-ionization capacitor C<sub>0</sub>; (b) Xenon Lamp<sub>8</sub>; (c) Main ionization inductor L<sub>9</sub>, respectively.

predicted trend follows closely along with the actual values until it reaches the failure threshold at degradation cycle 120. The estimated failure occurs at the degradation cycle 113, and the estimated RUP is 53 (113-60). The actual failure occurs at the degradation cycle 120, and the actual RUP is 60. Hence, the errors between the actual RUP and predicted RUP for C<sub>0</sub> is 7 degradation cycles.

To test different prediction lengths, which got close to failure gradually, as shown in Fig. 14 (a)-(c), the second group of prognosis experiments for C<sub>0</sub>, Lamp<sub>8</sub>, and L<sub>9</sub> are performed at degradation cycle 72 (120 × 60%), 81, 72, which is truncated at 60% of the testing run to failure data. In the same way, as shown in Fig. 15 (a)-(c), the third group of prognosis experiments for C<sub>0</sub>, Lamp<sub>8</sub>, and L<sub>9</sub> are performed at degradation cycle 84 (120 × 70%), 95, 84, which is truncated at 70% of the testing run to failure data. The estimated failure for C<sub>0</sub>, Lamp<sub>8</sub>, and L<sub>9</sub> occurs at the degradation cycle 118, 136, 123, and the actual failure occurs at degradation cycle 120, 135, 120. Hence, the errors between the actual RUP and predicted RUP are small, which are only 2, 1, and 3 degradation cycles for C<sub>0</sub>, Lamp<sub>8</sub>, and L<sub>9</sub>.

Fig. 16. (a)-(c) show the RUP estimation at every degradation cycle for C<sub>0</sub>, Lamp<sub>8</sub>, and L<sub>9</sub>, respectively. The blue line represents the RUP estimates, the estimated curve is relatively smooth and nearly linearity. Since the ideal degradation curve of the circuit component is linear, which proves the proposed method is effective and robust whenever the prognosis module is triggered.

4) COMPARED WITH THE STATE-OF-ART RUP ESTIMATION APPROACHES

The state-of-the-art methods, i.e., SVR [35], DCNN [41], LSTM [43], GRU [45], Bi-LSTM [44] are compared to the proposed method. The important parameters in these networks are summarized in Table 7. To illustrate that the Bi-GRU model is superior to SVR, DCNN, LSTM, GRU, and Bi-LSTM models, the average prediction error (Eave%) is used to evaluate the test results.

$$Eave\% = \frac{1}{M} \times \left| RUP_i - \widehat{RUP}_i \right| \times 100\% \quad (23)$$

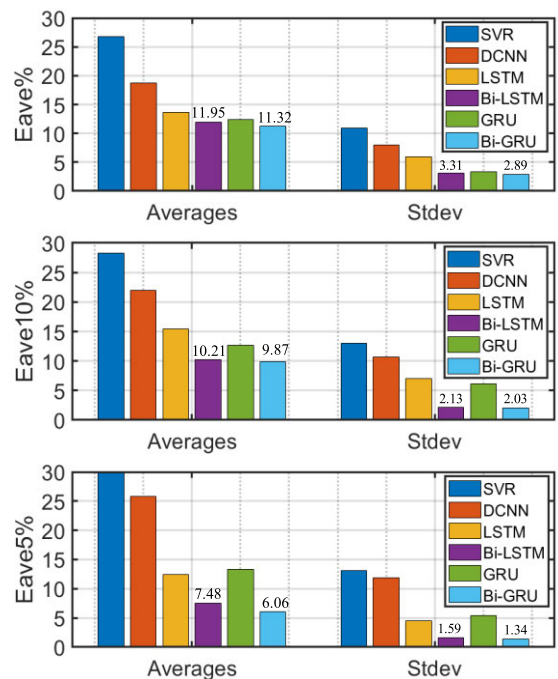


FIGURE 17. Compared Bi-GRU model with the SVR, DCNN, LSTM, GRU, and Bi-LSTM models.

where  $M$  denotes the number of samples in the degradation cycle,  $RUP_i$  and  $\widehat{RUP}_i$  represent actual RUP percentage and the predicted RUP percentage, respectively.

From the perspective of the engineering application, the RUP estimation accuracy of the final 10% and 5% of the whole degradation cycle (Eave10% & Eave5%) is essential. Therefore, the average prediction errors of Eave10% and Eave5% are calculated to validate the proposed approach. The results of existing methods are presented in Fig. 17. It reveals that the averages of Eave%, Eave10%, and Eave5% of the Bi-GRU model outperform the other compared methods.

More specifically, compared with SVR, the proposed method based on Bi-GRU solve the overfitting problem of shallow learning network caused by fluctuant noise features collected from multi-sensors. In addition, compared with

**TABLE 7. Parameter settings of different models.**

Networks	Parameter setting
SVR [35]	Maximum number of iterations was 100; The step of sliding window: 20.
DCNN [41]	Number of CNN cells: 5; Number of fully-connected layers: 1; Neurons in fully-connected layer: 100; The length of TW: 20; The batch size: 512. Epochs: 100. Optimizer: Adam.
LSTM [43]	Number of LSTM cell: 2. The number of standard feed-forward neural units: 2. The number of the feed-forward neural network: 2. The number of nodes: 8/8. 1-dimensional fully-connected layer.
GRU [45]	Number of GRU cells: 3; Number of fully-connected layers: 1; Numbers of neurons in the 3 layers: 64/32/16; The batch size: 10; The step of sliding window: 20; The epoch: 100; The learning rate: 0.01; The dropout rate: 0.2.
Bi-LSTM [44]	Number of LSTM cells: 3; Number of fully-connected layers: 1; Numbers of neurons in the 3 layers: 64/32/16; The batch size: 10; The step of sliding window: 20; The epoch: 100; The learning rate: 0.01; The dropout rate: 0.2.

SVR where predictions are based on the local degradation features, Bi-GRU can fuse multi-sensors data with variable time window sizes, which allows the neural network to be able to capture the short-range (local) and long-range (global) degradation features from multiple sensors.

DCNN can extract global and local features from some feature maps, but it generally requires large input parameters, which will inevitably be mixed with noise. On the contrary, Bi-GRU only requires fewer input parameters, which can shorten training time and control the balance between the noise and use information well. Furthermore, the network of Bi-GRU is more flexible in adjusting the structure parameters.

Compared with standard RNN methods (such as LSTM and GRU), Bi-LSTM and Bi-GRU can better represent in learning the hidden relationship between circuit components degradation features and RUP estimation. In addition, the amount of information extracted by the bidirectional approach is more than the single-direction approach. Especially at the late stage of the degradation process (i.e., Eave10% and Eave5%), the Bi-LSTM and Bi-GRU method can obtain better results.

For the Bi-LSTM model, the performance is similar to Bi-GRU in most of the cases. To further illustrate the superiority of the proposed Bi-GRU model, Fig. 17 shows that the averages of Eave%, Eave10% and Eave5% of Bi-GRU outperform Bi-LSTM by 0.63%, 0.34% and 1.42%, respectively. What's more, the standard deviation (Stdev) is calculated to explore the stability of the two models. It can be seen that the Stdevs of Eave%, Eave10%, and Eave5% of Bi-GRU outperform Bi-LSTM by 0.42%, 0.1%, and 0.25%, respectively. In summary, the experimental results show that the Bi-GRU achieves better RUP estimation performance than the above state-of-the-art methods.

## VI. CONCLUSION

In this paper, a novel prognostic framework is proposed by using a TSMIC algorithm for the selection of the most relevant features and Bi-GRU for RUP estimation, which

can reduce the fluctuation of the selected features as well as improve the accuracy of prediction. The contributions of this paper can be summarized as follows:

- Firstly, during the feature selection stage, owing to the lower sensitivity of other indicators to nonlinearity, a new feature selection metric, namely TSMIC, is proposed. The selected optimal feature set is as the input of the models, the prediction accuracy of all the prediction models are improved.
- Secondly, a Bi-GRU network model is employed to achieve RUP estimation. Several other state-of-the-art deep models, such as SVR, DCNN, LSTM, GRU, Bi-LSTM, are compared to prove the superiority of the proposed Bi-GRU. The test results indicate that the proposed method is superior to the existing models.

The future research aims to continue with the following goals: 1) A study how to extend RUP estimation in the case of multiple synchronous failures. 2) The deeper relations between features will be dug out with more distinctive approaches to reduce the error of RUP estimation further. 3) Some other deep models, such as the attention model, will be considered to enhance the robustness and generalization ability of circuit RUP estimation.

## REFERENCES

- [1] K. D. Humbird, J. L. Peterson, B. K. Spears, and R. G. McClarren, "Transfer learning to model inertial confinement fusion experiments," *IEEE Trans. Plasma Sci.*, vol. 48, no. 1, pp. 61–70, Jan. 2020.
- [2] E. I. Moses, "The national ignition facility and the national ignition campaign," *IEEE Trans. Plasma Sci.*, vol. 38, no. 4, pp. 684–689, Apr. 2010.
- [3] E. Zio, "Some challenges and opportunities in reliability engineering," *IEEE Trans. Rel.*, vol. 65, no. 4, pp. 1769–1782, Dec. 2016.
- [4] B. Sun, X. Jiang, K.-C. Yung, J. Fan, and M. G. Pecht, "A review of prognostic techniques for high-power white LEDs," *IEEE Trans. Power Electron.*, vol. 32, no. 8, pp. 6338–6362, Aug. 2017.
- [5] S. Zhao, V. Makis, S. Chen, and Y. Li, "Health assessment method for electronic components subject to condition monitoring and hard failure," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 138–150, Jan. 2019.
- [6] B. Cai, Y. Zhao, H. Liu, and M. Xie, "A data-driven fault diagnosis methodology in three-phase inverters for PMSM drive systems," *IEEE Trans. Power Electron.*, vol. 32, no. 7, pp. 5590–5600, Jul. 2017.
- [7] A. Alghassbi, S. Perinpanayagam, and M. Samie, "Stochastic RUL calculation enhanced with TDNN-based IGBT failure modeling," *IEEE Trans. Rel.*, vol. 65, no. 2, pp. 558–573, Jun. 2016.
- [8] B. Cai, X. Shao, Y. Liu, X. Kong, H. Wang, H. Xu, and W. Ge, "Remaining useful life estimation of structure systems under the influence of multiple causes: Subsea pipelines as a case study," *IEEE Trans. Ind. Electron.*, vol. 67, no. 7, pp. 5737–5747, Jul. 2020.
- [9] L. Ren, L. Zhao, S. Hong, S. Zhao, H. Wang, and L. Zhang, "Remaining useful life prediction for lithium-ion battery: A deep learning approach," *IEEE Access*, vol. 6, pp. 50587–50598, 2018.
- [10] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.
- [11] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, Jun. 2015.
- [12] D. Binu and B. S. Kariyappa, "RideNN: A new rider optimization algorithm-based neural network for fault diagnosis in analog circuits," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 2–26, Jan. 2019.
- [13] C. Zhang, Y. He, L. Yuan, and S. Xiang, "Analog circuit incipient fault diagnosis method using DBN based features extraction," *IEEE Access*, vol. 6, pp. 23053–23064, 2018.



- [14] A. S. S. Vasan, B. Long, and M. Pecht, "Diagnostics and prognostics method for analog electronic circuits," *IEEE Trans. Ind. Electron.*, vol. 60, no. 11, pp. 5277–5291, Nov. 2013.
- [15] B. Long, S. Tian, and H. Wang, "Diagnostics of filtered analog circuits with tolerance based on LS-SVM using frequency features," *J. Electron. Test.*, vol. 28, no. 3, pp. 291–300, Jun. 2012.
- [16] C. Zhang, Y. He, L. Yuan, and F. Deng, "A novel approach for analog circuit fault prognostics based on improved RVM," *J. Electron. Test.*, vol. 30, no. 3, pp. 343–356, Jun. 2014.
- [17] Z. Liu, T. Liu, J. Han, S. Bu, X. Tang, and M. Pecht, "Signal model-based fault coding for diagnostics and prognostics of analog electronic circuits," *IEEE Trans. Ind. Electron.*, vol. 64, no. 1, pp. 605–614, Jan. 2017.
- [18] L. Liao and F. Köttig, "A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction," *Appl. Soft Comput.*, vol. 44, pp. 191–199, Jul. 2016.
- [19] Z. Jia, Z. Liu, C.-M. Vong, and X. Liu, "Real-time response-based fault analysis and prognostics techniques of nonisolated DC-DC converters," *IEEE Access*, vol. 7, pp. 67996–68009, 2019.
- [20] S. Mahmoodi, S. H. Ranjesh, and Y. Q. Zhao, "Condition-based maintenance policies for a multi-unit deteriorating system subject to shocks in a semi-Markov operating environment," *Qual. Eng.*, pp. 1–12, Mar. 2020.
- [21] J. Ben Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, and F. Fnaiech, "Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network," *Mech. Syst. Signal Process.*, vols. 56–57, pp. 150–172, May 2015.
- [22] M. Li, W. Xian, B. Long, and H. Wang, "Prognostics of analog filters based on particle filters using frequency features," *J. Electron. Test.*, vol. 29, no. 4, pp. 567–584, Aug. 2013.
- [23] J. Z. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1803–1836, Jul. 2011.
- [24] Y. Qian and R. Yan, "Remaining useful life prediction of rolling bearings using an enhanced particle filter," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 10, pp. 2696–2707, Oct. 2015.
- [25] M. Ibrahim, N. Y. Steiner, S. Jemei, and D. Hissel, "Wavelet-based approach for online fuel cell remaining useful lifetime prediction," *IEEE Trans. Ind. Electron.*, vol. 63, no. 8, pp. 5057–5068, Aug. 2016.
- [26] J. P. Kharoufeh and S. M. Cox, "Stochastic models for degradation-based reliability," *IIE Trans.*, vol. 37, no. 6, pp. 533–542, Jun. 2005.
- [27] J. R. Celaya, A. Saxena, C. S. Kulkarni, S. Saha, and K. Goebel, "Prognostics approach for power MOSFET under thermal-stress aging," in *Proc. Annu. Rel. Maintainability Symp.*, Reno, NV, USA, Jan. 2012, pp. 1–6.
- [28] L. Yuan, Y. He, J. Huang, and Y. Sun, "A new neural-network-based fault diagnosis approach for analog circuits by using kurtosis and entropy as a preprocessor," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 3, pp. 586–595, Mar. 2010.
- [29] W. He, Y. He, and B. Li, "Generative adversarial networks with comprehensive wavelet feature for fault diagnosis of analog circuits," *IEEE Trans. Instrum. Meas.*, early access, Jan. 24, 2020, doi: 10.1109/TIM.2020.2969008.
- [30] Y. Tan, Y. Sun, and X. Yin, "Analog fault diagnosis using S-transform preprocessor and a QNN classifier," *Measurement*, vol. 46, no. 7, pp. 2174–2183, Aug. 2013.
- [31] W. He, Y. He, B. Li, and C. Zhang, "A Naive-Bayes-based fault diagnosis approach for analog circuit by using image-oriented feature extraction and selection technique," *IEEE Access*, vol. 8, no. 3, pp. 5065–5079, 2019.
- [32] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, "Enabling health monitoring approach based on vibration data for accurate prognostics," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 647–656, Jan. 2015.
- [33] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, Dec. 2011.
- [34] L. Yan, H. Wang, Z. Liu, J. Zhou, and B. Long, "A novel noise-assisted prognostic method for linear analog circuits," *J. Electron. Test.*, vol. 33, no. 2, pp. 559–572, Oct. 2017.
- [35] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, "Remaining useful life estimation based on nonlinear feature reduction and support vector regression," *Eng. Appl. Artif. Intell.*, vol. 26, no. 7, pp. 1751–1760, Aug. 2013.
- [36] X. Zheng and H. Fang, "An integrated unscented Kalman filter and relevance vector regression approach for lithium-ion battery remaining useful life and short-term capacity prediction," *Rel. Eng. Syst. Saf.*, vol. 144, pp. 74–82, Dec. 2015.
- [37] B. Cai, Y. Liu, Z. Liu, Y. Chang, and L. Jiang, "Application of Bayesian networks in reliability evaluation," in *Bayesian Networks for Reliability Engineering*, 2020, pp. 1–25.
- [38] B. Cai, Y. Liu, Z. Liu, Y. Chang, and L. Jiang, "Reliability evaluation methodology of complex systems based on dynamic object-oriented Bayesian networks," in *Bayesian Networks for Reliability Engineering*, 2020, pp. 87–107.
- [39] B. Cai, Y. Liu, Q. Fan, Y. Zhang, Z. Liu, S. Yu, and R. Ji, "Multi-source information fusion based fault diagnosis of ground-source heat pump using Bayesian network," *Appl. Energy*, vol. 114, pp. 1–9, Feb. 2014.
- [40] B. Cai, Y. Liu, Z. Liu, X. Tian, X. Dong, and S. Yu, "Using Bayesian networks in reliability evaluation for subsea blowout preventer control system," *Rel. Eng. Syst. Saf.*, vol. 108, pp. 32–41, Dec. 2012.
- [41] L. Ren, Y. Sun, H. Wang, and L. Zhang, "Prediction of bearing remaining useful life with deep convolution neural network," *IEEE Access*, vol. 6, pp. 13041–13049, 2018.
- [42] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [43] P. Malhotra, V. Tv, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "Multi-sensor prognostics using an unsupervised health index based on LSTM encoder-decoder," *Comput. Sci.*, 2016.
- [44] C.-G. Huang, H.-Z. Huang, and Y.-F. Li, "A bi-directional LSTM prognostics method under multiple operational conditions," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8792–8802, Nov. 2019.
- [45] L. Ren, X. Cheng, X. Wang, J. Cui, and L. Zhang, "Multi-scale dense gate recurrent unit networks for bearing remaining useful life prediction," *Future Gener. Comput. Syst.*, vol. 94, pp. 601–609, May 2019.
- [46] J. Yan, M. Koç, and J. Lee, "A prognostic algorithm for machine performance assessment and its application," *Prod. Planning Control*, vol. 15, no. 8, pp. 796–801, Dec. 2004.
- [47] J. Huang, X. Hu, and F. Yang, "Support vector machine with genetic algorithm for machinery fault diagnosis of high voltage circuit breaker," *Measurement*, vol. 44, no. 6, pp. 1018–1027, Jul. 2011.
- [48] M. Aminian and F. Aminian, "A modular fault-diagnostic system for analog electronic circuits using neural networks with wavelet transform as a preprocessor," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 5, pp. 1546–1554, Oct. 2007.
- [49] C. Yang, S. Tian, B. Long, and F. Chen, "Methods of handling the tolerance and test-point selection problem for analog-circuit fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 1, pp. 176–185, Jan. 2011.
- [50] C. Alippi, M. Catelani, A. Fort, and M. Mugnaini, "SBT soft fault diagnosis in analog electronic circuits: A sensitivity-based approach by randomized algorithms," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 5, pp. 1116–1125, Oct. 2002.



**BOLUN DU** received the M.Sc. degree from Hunan University, Changsha, China, in 2018. He is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Automation, Wuhan University, Wuhan, China. His research interests include remaining useful life prediction, condition monitoring, fault diagnosis, machine learning, and signal processing.

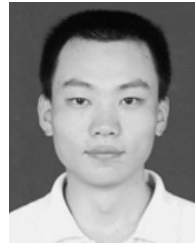


**YIGANG HE** (Member, IEEE) received the Ph.D. degree in electrical engineering from Xian Jiaotong University, Xi'an, China, in 1996. He was a Senior Visiting Scholar with the University of Hertfordshire, Hatfield, U.K., in 2002. From 2011 to 2017, he has worked as the Head of the School of Electrical Engineering and Automation, Hefei University of Technology. In December 2017, he joined Wuhan University, China, where he currently works as the Vice-Head of the

School of Electrical Engineering and Automation. His research interests include power electronic circuit theory and its applications, testing and fault diagnosis of analog and mixed-signal circuits, electrical signal detection, smart grid, satellite communication monitoring, and intelligent signal processing. He has authored or coauthored 300 journal articles and conference papers, which was included more than 1000 times in Science Citation Index of American Institute for Scientific Information in the aforementioned areas. He was a winner of the National Science Fund for Distinguished Young Scholars, in 2009.



**BAORAN AN** received the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, in 2014. He is currently an Associate Research Fellow with the Institute of Computer Application, China Academy of Engineering Physics. He has been the Chief Expert on IPD projects, since 2017. His current research interests include fault diagnostics and prognostics of analog and mixed-signal circuits, and battery capacity prognostic.



**CHAOLONG ZHANG** received the Ph.D. degree from the Hefei University of Technology, in 2018. He is currently a Postdoctoral Researcher of electrical engineering and automation with Wuhan University. He is also an Associate Professor with the School of Physics and Electronic Engineering, Anqing Normal University. His current research interests include fault diagnostics and prognostics of analog and mixed-signal circuits, battery capacity prognostic, satellite communication monitoring, and intelligent signal processing.

...