Energy Engineering

Tech Science Press

Doi:10.32604/ee.2025.075298

ARTICLE

Effective Prediction of Aging and Remaining Useful Life of Proton Exchange Membrane Fuel Cell via Kolmogorov-Arnold Network Based Gated Recurrent Unit

Wenqiang Xie*, Xiaolong Xiao, Fangfang Zhu, Ziran Guo and Xiaoxing Lu

China State Grid Jiangsu Electric Power Co. Ltd. Research Institute, Nanjing, 211103, China

*Corresponding Author: Wenqiang Xie. Email: wenqiangxie1@outlook.com

Received: 29 October 2025; Accepted: 18 December 2025; Published: 24 September 2026

ABSTRACT: In the framework of the comprehensive energy transition, the proton exchange membrane fuel cell (PEMFC) powered by renewable energy emerges as a promising alternative, particularly with relevance to applications like electric vehicles (EVs), where clean and efficient power sources are crucial. However, the accurate prediction of PEMFC performance degradation poses significant challenges due to the combined effects of complex and variable aging mechanisms and operational conditions, which are especially critical in the context of EV applications where reliability and durability directly impact vehicle performance and user safety. These challenges pose notable constraints on the feasibility of its large-scale commercialization in the automotive sector. To address these issues, this study proposes an efficient prediction method, namely the Kolmogorov-Arnold network-based gated recurrent unit (GRU-KAN), aiming to improve prediction accuracy and optimize maintenance management strategies for PEMFC systems in EVs. Firstly, median filtering is employed to mitigate the interference from noise and extreme environmental factors. Furthermore, Pearson correlation analysis is utilized to screen out parameters strongly correlated with the output. Lastly, GRU-KAN is applied to predict the aging trend and remaining useful life (RUL). To validate the superiority of GRU-KAN, aging data from different operating states are employed. Simulation results demonstrate that GRU-KAN exhibits higher prediction accuracy compared to other methods. For instance, under a dynamic state and with a 50% training set, the root-mean-square error of aging prediction errors for GRU-KAN is reduced by 77.08% compared to GRU, while also providing the most accurate RUL prediction results, thereby enhancing the prognostic capability for PEMFC health management in electric vehicles.

KEYWORDS: Proton exchange membrane fuel cell; gated recurrent unit; Kolmogorov-Arnold network; aging; remaining useful life

1 Introduction

In contemporary society, there exists a pervasive adoption of renewable energy sources which are widely regarded as the cornerstones for the establishment of a sustainable future [1,2]. Among the various clean energy alternatives, hydrogen energy stands out as a luminous novelty, capturing considerable attention as a preferred option [3,4]. As a clean and highly efficient energy carrier, hydrogen holds immense potential to revolutionize the global energy landscape [5,6]. Hydrogen is capable of directly converting chemical energy into electrical energy through chemical reactions within fuel cells, with water being the sole byproduct, thereby rendering it an ideal choice for achieving pollution-free power generation [7]. In exploring the immense potential of hydrogen energy as a future clean energy source, proton exchange membrane fuel cell

(PEMFC) emerges as an indispensable cornerstone in the realm of hydrogen applications [8]. Leveraging its advanced technological prowess, PEMFC paves a green and feasible avenue toward the realization of zero-emission transportation systems and distributed energy networks [9,10]. However, the intrinsic construction and material characteristics of PEMFC, coupled with the operational environment, can result in irreversible performance degradation [11]. Poor durability and high costs continue to pose significant barriers to the global commercialization of PEMFC. Accurate prediction of PEMFC aging trends and estimation of its remaining useful life (RUL) holds immense significance in minimizing maintenance costs, averting catastrophic failures, and prolonging the operational lifespan of fuel cells [12,13]. Therefore, conducting in-depth research into the aging mechanisms of PEMFC and enhancing the accuracy of aging prediction represent urgent and critical tasks in the current fuel cell domain [14]. PEMFC aging prediction methods are generally categorized into three types: model-driven method, data-driven method, and hybrid method.

Despite the advancements in PEMFC technology, the accurate prediction of its aging and RUL remains a significant challenge. Existing methods have shown limitations in handling the complex and variable aging mechanisms and operational conditions of PEMFC. To address these challenges, this study proposes an innovative prediction method, namely the Kolmogorov-Arnold network based gated recurrent unit (GRU-KAN). This method combines the strengths of GRU in capturing temporal dependencies and KAN's ability to model complex nonlinear relationships, aiming to provide a more accurate and reliable solution for PEMFC aging and RUL prediction. The proposed approach not only enhances the prediction accuracy but also offers a robust framework for optimizing maintenance management strategies, thereby providing a valuable tool that can contribute to the development of more reliable PEMFC systems, with potential applications in areas such as electric vehicles.

The model-driven method uses a mathematical model to characterize the degradation process of PEMFC and then extracts the parameters of the model through advanced filter technology. In particular, there are various models to characterize the aging characteristics of PEMFC, including the mechanism model [15], empirical model [16], and semi-empirical model [17]. In addition, for the aging process in different physical fields, some researchers have also innovatively proposed improved models to further improve the prediction accuracy and practicability. Zhou et al. [14] have proposed a multi-physical aging model, which is rooted in the intricate dynamics of PEMFC reactivity loss and reactant mass transfer loss. Zhu and Chen [18] have introduced a novel dynamic degradation model for PEMFC, grounded in the Gaussian process state space framework, offering enhanced flexibility in addressing the intricate degradation processes and inherent uncertainties associated with PEMFC. However, the reliance of model-driven methods on the precision of the model, coupled with their restricted versatility across diverse operating conditions, undermines the applicability of this prediction approach.

The data-driven method stands as a powerful approach, shunning the need for intricate explorations into the complex mechanisms behind fuel cell decay. Instead, it harnesses statistical principles, cutting-edge machine learning techniques, and sophisticated artificial intelligence to uncover hidden patterns within vast experimental datasets. Currently, various approaches are applied to the issue of PEMFC aging prediction, including the extreme learning machine (ELM) [19], long short-term memory (LSTM) [20], and recurrent neural network (RNN) [21]. Zheng et al. [20] have developed an adaptive prediction strategy that involves extracting health indicators by identifying a series of linear parameter-varying models and employing an ensemble echo state network for making predictions. An automatic prediction method for the adaptive neuro-fuzzy inference system with fuzzy c-means by using particle swarm optimization (PSO) has been established by Liu et al. [22], and the best short-term prediction performance was obtained by effectively adjusting the parameters of PSO. Based on multiple key control parameters of batteries, Chen et al. [23] have employed a grey neural network combined with cuckoo search algorithm to predict degradation under different operating conditions, aiming to ensure the long-term effective operation of PEMFC. Zhou et al. [24] have proposed a hybrid autoregressive and moving average-time delay neural network, which is applied to capture the linear and nonlinear characteristics of battery data, thereby improving the accuracy of aging predictions.

The hybrid method masterfully blends the strengths of both model-driven and data-driven approaches, boosting the precision and reliability of RUL prediction. Yet, this method remains inherently tied to the accuracy of its foundational model. Challenges arise from the diverse aging mechanisms, environmental disruptions, and unpredictable noise in measurements, all of which can undermine the accuracy of the model. Consequently, the hybrid method still grapples with similar constraints as the model-driven method in certain realms [25].

In conclusion, among the aforementioned three prediction methods, only the data-driven approach does not require model building, thereby effectively circumventing the impact of aging factors on model establishment. However, current scholarly investigations into data-driven methodologies demonstrate a disproportionate emphasis on direct prediction using battery performance metrics, overlooking the impacts of stochastic noise and transient environmental perturbations on measurement integrity. Furthermore, some studies, like studies [22,23], employ heuristic algorithms to continuously adjust neural network parameters to achieve prediction accuracy. Nevertheless, heuristic algorithms may get trapped in local optimal solutions and are highly sensitive to parameters where minor changes can lead to significant alterations in prediction performance. Considering the notable gaps identified in the existing literature, this study introduces an effective prediction method, namely Kolmogorov-Arnold network based gated recurrent unit (GRU-KAN), aiming to elevate the precision, reliability, and practical applicability of prediction related to equipment aging trends and RUL. The main contributions can be summarized as follows:

- Aiming at the noise interference and parameter changes caused by aging, median filtering technology is employed to smooth the original data, effectively reducing the impact of noise. Subsequently, Pearson correlation analysis (PCA) is applied to extract the correlation between parameters, providing a more accurate and reliable data foundation for subsequent prediction models.
- By integrating the flexible activation functions of the Kolmogorov-Arnold network to learn complex nonlinear relationships, along with the advantages of the GRU in capturing long-term dependencies in time-series data, GRU-KAN is harnessed for predicting the aging trends and RUL of PEMFC.

Under steady and dynamic operating conditions, comprehensive case studies complemented by sensitivity analysis are developed to compare GRU-KAN against other competitive prediction methods, demonstrating exceptional performance in prediction accuracy, computational efficiency, and model stability of GRU-KAN.

2 Aging Factors and Evaluation of PEMFC

This section explores the working principle of PEMFC, analyzes the impact of aging factors on its performance, and introduces the key performance parameters of PEMFC.

2.1 Overview of PEMFC

PEMFC is widely adopted for its clean, green, and effective characteristics. Inside the cell, hydrogen is skillfully decomposed into protons and electrons at the anode. Specifically, the anode reaction involves the oxidation of hydrogen gas, which can be represented by the equation:

$$2H_2 \rightarrow 4H^+ + 4e^- \tag{1}$$

These protons travel through the proton exchange membrane (PEM) to the cathode, while the electrons flow through the external circuit, producing electricity. At the cathode, the protons combine with oxygen gas and the electrons to form water, with the cathode reaction represented by:

$$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O \tag{2}$$

The complete overall reaction of the PEMFC is obtained by combining the anode and cathode reactions, where the protons and electrons are balanced, resulting in the equation:

$$2H_2 + O_2 \rightarrow 2H_2O \tag{3}$$

This process generates water and heat, thereby converting chemical energy into electrical energy efficiently and directly.

2.2 Aging Factors

The PEMFC single cell boasts a sophisticated core composition, encompassing PEM, catalyst layer, diffusion layer, flow field, and sealing section [26]. Each of these components plays a distinct role in the operation of PEMFC, driving the efficient conversion of energy [27]. Yet, as time progresses, every component undergoes varying degrees of aging, a natural phenomenon that silently erodes the efficiency and performance of PEMFC. As vividly depicted in Fig. 1, it is easy to discern the intricate interplay of key factors influencing PEMFC efficiency and the underlying aging mechanisms [28]. These mechanisms stem from a multitude of unpredictable and volatile environmental factors, creating a relatively limited scope for regulation. However, a profound understanding of these aging mechanisms will significantly enhance our ability to predict the aging process and RUL of PEMFC with greater accuracy [29].

2.3 Remaining Useful Life

In the field of engineering, RUL is usually assessed based on a variety of factors such as physical wear, chemical corrosion, and environmental impact of the equipment [30]. Various predictive models and analysis tools are utilized to estimate the RUL of the equipment to schedule repairs, replacements, or upgrades promptly. RUL used in engineering can be characterized as follows [31]:

$$RUL = T_{\text{failure}} - T_{\text{BOL}} \tag{4}$$

where T_{BOL} represents the current point in time when the judgment function is taken; T_{failure} is the time when the cell reaches the failure threshold; RUL indicates the amount of time or use period in which normal work can also continue in the current state. Note that T_{BOL} is set to 550 h in this study.

The value of T_{BOL} is set to 550 h based on multiple factors. Firstly, this value is determined according to the manufacturer's technical specifications and the results of accelerated aging tests, which indicate that after 550 h of operation, the PEMFC begins to exhibit noticeable signs of performance degradation under typical operating conditions. Secondly, it aligns with the industry standard for evaluating the durability of PEMFC systems, providing a consistent benchmark for comparing the performance and lifespan of different PEMFC units. Additionally, this threshold ensures a balance between sufficient data collection for accurate RUL prediction and the practical operational lifespan expectations of PEMFC in real-world applications.

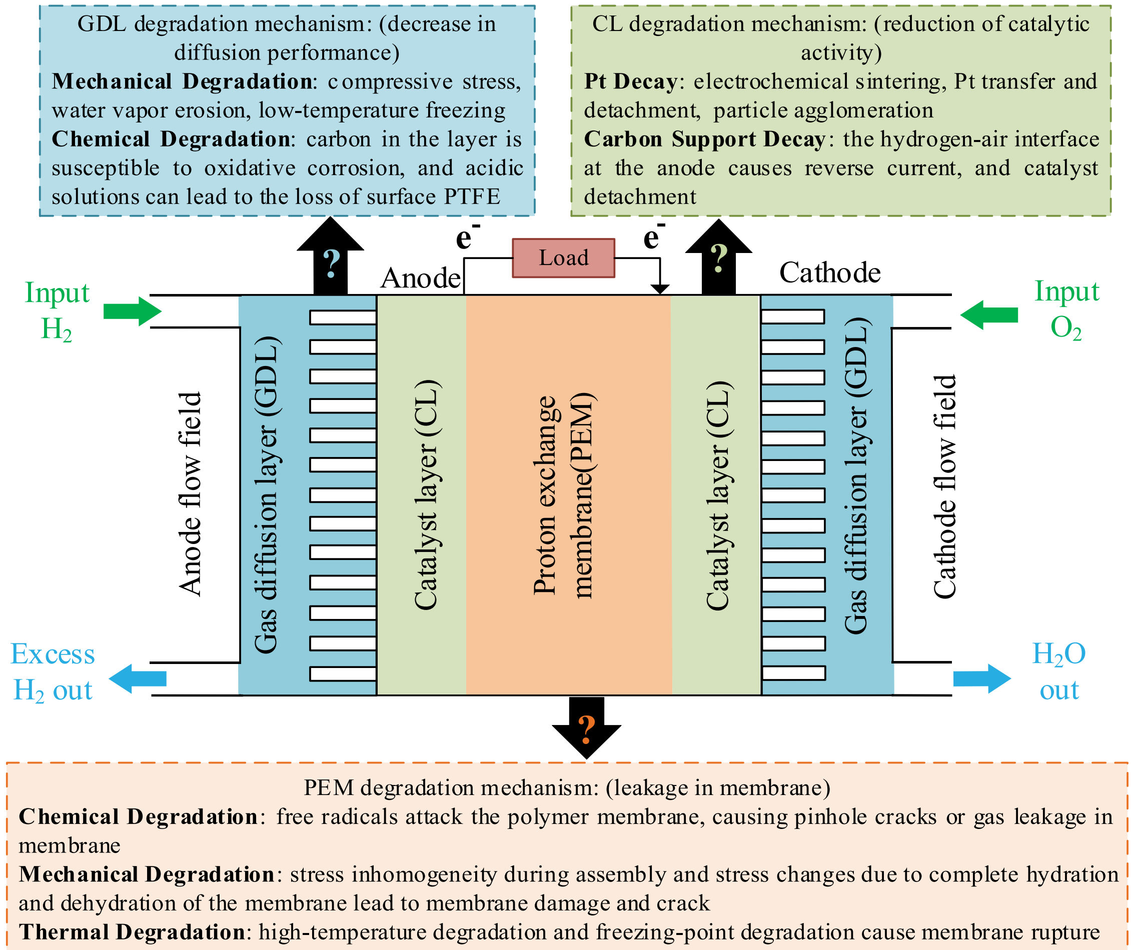


Figure 1: Major degradation mechanism and principal structure of PEMFC

3 Methodology

In this section, the effective prediction method GRU-KAN for aging and RUL is introduced.

3.1 Pearson Correlation Analysis

Correlation analysis is a statistical technique used to evaluate the relationship between two or more variables. Through correlation analysis, it is possible to determine whether there is a relationship between these variables and the strength and direction of this relationship. PCA was chosen for feature selection in this study due to several considerations. Firstly, in the initial exploratory analysis of the PEMFC aging data, it was observed that many of the key features exhibited approximate linear relationships with the output variables. This suggested that a linear correlation measure like Pearson's could effectively identify the features with significant influence on the prediction target. Secondly, the computational efficiency of Pearson correlation analysis is relatively high, making it suitable for processing the large-scale datasets commonly encountered in PEMFC aging studies. Although more robust nonlinear feature selection methods exist, such as kernel-based methods and mutual information, these often come with higher computational complexity and require more careful parameter tuning. Preliminary experiments found that Pearson correlation analysis was sufficient to meet the feature screening requirements of this study. To ensure the comprehensiveness of the feature selection process, a comparison was also conducted with a nonlinear method, namely the mutual information method. The results indicated that, for the specific characteristics of the PEMFC aging data in this study, Pearson correlation analysis and mutual information yielded similar feature selection results. Given the simplicity and efficiency of Pearson's method, it was ultimately adopted as the primary feature selection approach. PCA is used to measure the linear relationship between variable data, quantifying the strength and direction of this relationship by calculating the Pearson correlation coefficient [32].

PCA is extensively applied to measure the correlation between two variables, especially in the analysis of complex data of multi-factor comprehensive influence, which has high reliability [33]. In PCA, for two variables x and y, several groups of data can be obtained through experiments, denoted by (x_i, y_i) $(i = 1, 2, \ldots, n)$. PCA is defined as the quotient of the covariance and standard deviation of the estimated sample, and the mathematical expression of the correlation coefficient r is

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 \sum_{i=1}^{n} (y_i - \overline{y})^2}} \tag{5}$$

where $\overline{x}$ and $\overline{y}$ are the mean values of n test values, respectively.

The value of r is between −1 and 1. If $r > 0$, indicating that the two variables are positively correlated; otherwise, it indicates that the two variables are negatively correlated, and the greater the absolute value of r, the stronger the correlation. PCA is roughly divided into three types of correlation strength: the range of weak correlation is [0.1, 0.3], the range of medium correlation is [0.3, 0.5], and the range of strong correlation is [0.5, 1].

3.2 Gated Recurrent Unit

GRU, proposed by Chung et al. in 2014 [34], is an optimized recurrent neural network architecture, and its structure diagram is shown in Fig. 2. By introducing two gating mechanisms, namely reset gate and update gate, GRU precisely regulates the flow of information, addressing the gradient vanishing problem faced by traditional RNN when processing long sequences. Specifically, the reset gate determines the degree of influence of the previous hidden state on the current hidden state, while the update gate determines the impact of the current hidden state on the next hidden state. This design not only effectively updates short-term dependent information but also appropriately resets long-term dependent information, making GRU characterized by parameter parsimony, rapid training, and strong short-term dependency capturing capabilities [35].

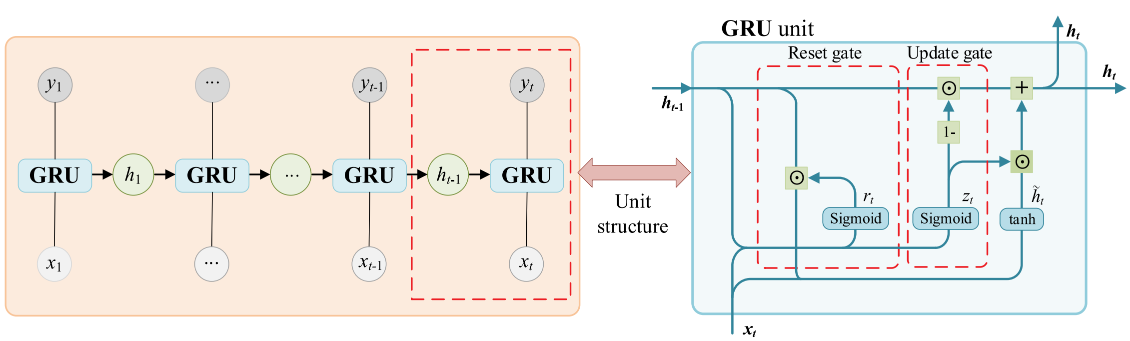


Figure 2: GRU network structure diagram

The GRU network structure comprises several key components that work together to process sequential data and capture temporal dependencies. These components include:

(1) Reset gate: It determines how much of the previously hidden state information is forgotten, aiding in capturing short-term dependencies. At time step t, the reset gate r_t is calculated as follows:

$$r_t = \sigma\left(W_{\mathrm{r}} \cdot \left[\mathrm{h}_{t-1}, x_t\right] + b_{\mathrm{r}}\right) \tag{6}$$

where σ denotes the Sigmoid activation function; W_{r} is the weight matrix of the reset gate; h_{t-1} represents the hidden state at the previous moment; x_t indicates the input at time t; b_{r} means the bias of the reset gate.

(2) Update gate: It determines how much new information is added to the current state, facilitating the capture of long-term dependencies, which is defined as [34]:

$$z_t = \sigma\left(W_{\mathrm{z}} \cdot \left[\mathrm{h}_{t-1}, x_t\right] + b_{\mathrm{z}}\right) \tag{7}$$

where W_{z} and b_{z} represent the weight matrix and bias of the update gate, respectively.

(3) Candidate hidden state: It is an intermediate state that is constructed based on the current input and the hidden state from the previous time step, providing a candidate for the subsequent hidden state update, which is defined as:

$$\tilde{\mathrm{h}}_t = \tanh\left(W_{\mathrm{q}} \cdot \left[r_t \ast \mathrm{h}_{t-1}, x_t\right] + b_{\mathrm{q}}\right) \tag{8}$$

where W_{q} and b_{q} denote the weight matrix and bias associated with the candidate hidden state, respectively.

(4) Final hidden state: The final hidden state, which is the ultimate output of GRU, is computed by combining the update gate, the hidden state from the previous time step, and the candidate hidden state, representing the network's memory at the current time step.

$$\mathrm{h}_t = \left(1 - z_t\right) \ast \mathrm{h}_{t-1} + z_t \ast \tilde{\mathrm{h}}_t \tag{9}$$

3.3 Kolmogorov-Arnold Network

In stark contrast to multi-layer perceptron (MLP) neural networks that rely on traditional fixed activation functions, Kolmogorov-Arnold network (KAN) is profoundly inspired by the Kolmogorov-Arnold theorem [36]. It ingeniously revolutionizes the weight parameters on the edges of MLP into univariate functions, parameterized in the form of flexible activation functions. As depicted in Fig. 3, the network structures of MLP and KAN present a striking contrast.

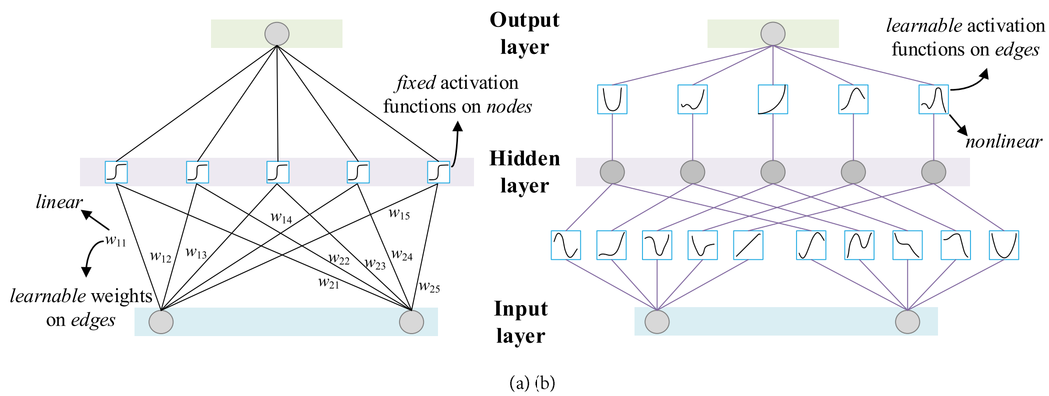


Figure 3: Comparison of neural network structures: (**a**) MLP and (**b**) KAN

A simple two-layer KAN network architecture KAN can be outlined as [36]:

$$f(x_1,\ldots,x_n) = \sum_{q=1}^{2n+1} \Phi_q\left(\sum_{p=1}^{n} \phi_{q,p}(x_p)\right) \quad (10)$$

where $\phi_{q,p}$ denotes a univariate function representing the internal function, which processes the pth component of the input vector x and contributes the result to the qth external function; Φ_q represents a univariate function for the external function, which sums the outputs of the internal function layer to form the final output of the model. In order to enable KAN to maintain higher accuracy while preserving flexibility, Liu et al. in 2024 [36] proposed a residual activation strategy. By introducing the basis function $b(x)$, $\phi(x)$ is the sum of $b(x)$ and the spline function $\mathrm{spline}(x)$, which is calculated as:

$$\phi(x) = w_b \cdot b(x) + w_s \cdot \mathrm{spline}(x) \quad (11)$$

where w_b and w_s indicate the learnable parameters. The basis function $b(x)$ can be defined as:

$$b(x) = \mathrm{SiLU}(x) = x/(1 + e^{-x}) \quad (12)$$

where $b(x)$ is set as the SiLU activation function. $\mathrm{spline}(x)$ is represented as a linear combination of B-splines, parameterized as follows [36]:

$$\mathrm{spline}(x) = \sum_i c_i B_i(x) \quad (13)$$

where c_i denotes the trainable parameter; $B_i(x)$ means a predefined B-spline basis function. During the training process, the spline parameter c_i is continuously optimized to adjust the shape of the B-spline, thereby enabling it to better fit the training data.

3.4 GRU-KAN

GRU is known for its computational efficiency and strong ability to process short-term temporal data due to its simplified design. However, GRU has limitations in capturing long-range dependencies, especially in complex situations that require recognizing patterns at multiple scales. When predicting PEMFC aging, accurately identifying slow-changing trends along with short-term variations is crucial for understanding electrochemical degradation.

The integration of KAN into the GRU architecture is motivated by the need to better capture the complex, non-linear degradation dynamics of PEMFCs. Traditional GRUs employ fixed activation functions, which possess limited flexibility in modeling intricate, physics-informed non-linearities inherent in voltage decay curves. In contrast, KAN replaces these static functions with learnable, univariate functions parameterized by B-splines, allowing the network to adaptively adjust activation function shapes during training. This adaptive mechanism enables the model to discover and represent underlying multi-scale degradation patterns more effectively than predefined functions, thereby enhancing GRU's capability in capturing both short-term fluctuations and long-term trends in the PEMFC aging process.

Specifically, in the proposed GRU-KAN model, the KAN structure is employed to replace the standard activation functions within the GRU cell, particularly in the calculation of the candidate hidden state, where capturing complex transformations is most critical for learning temporal dynamics. Through this hybrid strategy with optimized feature transformations, the network achieves improved processing of sequential data and better representation of information across different temporal scales. The network structure of the GRU-KAN model is illustrated in Fig. 4.

To elucidate the architectural differences and validate the implementation, the pseudocode for the core training procedures of RNN-KAN and GRU-KAN is provided in Tables 1 and 2, respectively.

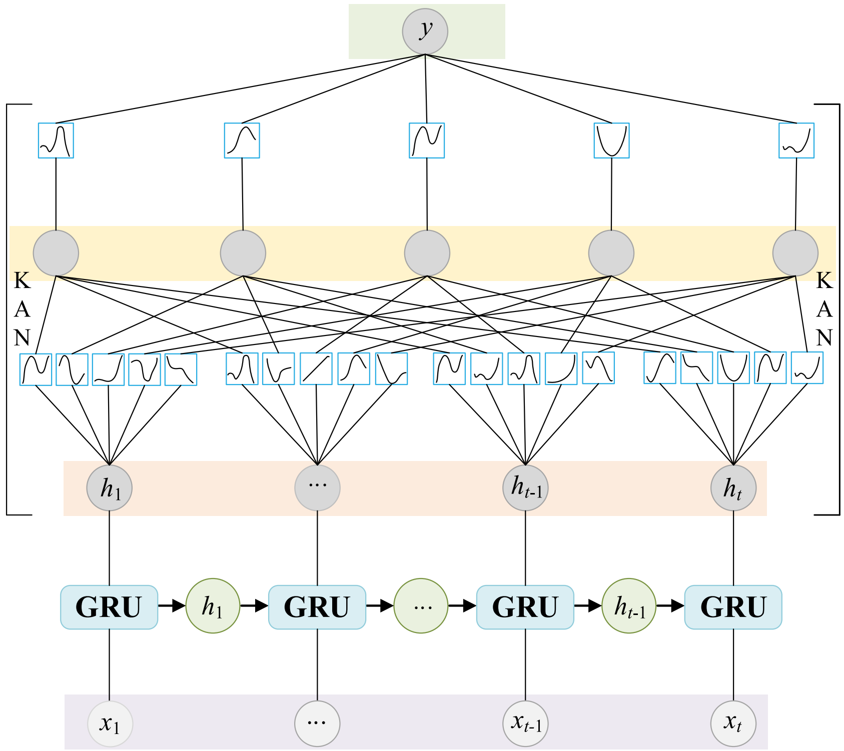


Figure 4: The network structure diagram of GRU-KAN

Table 1: Pseudocode for RNN-KAN algorithm

Step	Process
1	Input: Sequence data X, hidden dimension h, output dimension o
2	Initialize: RNN weights W_{hh}, W_{xh}, b_h; KAN parameters c_i, w_b, w_s
3	for epoch = 1 to N_{epochs} do
4	for each batch (x_b, y_b) in training data do
5	$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_b + b_h)$
6	$y_{\text{pred}} = \text{KAN}(h_t)$
7	$L = \text{MSE}\left(y_b, y_{\text{pred}}\right)$
8	Compute gradients ∇L
9	Update parameters via Adam optimizer
10	end for
11	end for
12	Output: Predicted sequence Y_{pred}

Table 2: Pseudocode of GRU-KAN algorithm

Step	Process
1	Input: Sequence data X, hidden dimension h, output dimension o
2	Initialize: GRU weights $W_z, W_r, W_q, b_z, b_r, b_q$; KAN parameters

(Continued)

Table 2 (continued)

Step	Process
3	for epoch = 1 to N_{epochs} do
4	for each batch (x_b, y_b) in training data do
5	$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_b] + b_z\right)$
6	$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_b] + b_r\right)$
7	$\tilde{h}_t = \text{KAN}\left(r_t \odot h_{t-1}, x_b\right)$
8	$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$
9	$y_{\text{pred}} = W_o \cdot h_t + b_o$
10	$L = \text{MSE}\left(y_b, y_{\text{pred}}\right)$
11	Compute gradients ∇L
12	Update parameters via Adam optimizer
13	end for
14	end for
15	Output: Predicted sequence Y_{pred}

The primary distinction between the two algorithms lies in their recurrent mechanisms and KAN integration. RNN-KAN employs a standard recurrent cell followed by a KAN layer for output transformation. In contrast, GRU-KAN leverages gating mechanisms and integrates KAN directly into the candidate hidden state computation, enabling adaptive activation learning. This architectural difference allows GRU-KAN to better capture multi-scale degradation patterns, contributing to its superior performance in PEMFC aging prediction under dynamic conditions.

3.5 Effective Prediction Strategy

In the research focused on PEMFC aging and RUL prediction based on data-driven methods, the key lies in conducting an in-depth analysis of vast amounts of test data, with the aim of accurately and effectively capturing and predicting the aging trends of cells. This study introduces an innovative PEMFC aging and RUL prediction strategy, with its core process illustrated in Fig. 5. Firstly, the pre-processed data is input into the GRU-KAN model to predict the voltage decay curve. The GRU-KAN model leverages its unique architecture to handle the temporal structure of the input data. The GRU component captures the short-term dependencies in the data through its reset and update gates, which regulate the flow of information and help in extracting local features. Meanwhile, the KAN component interacts with the temporal structure by learning the complex nonlinear relationships across different time scales. Its adaptive activation functions enable the model to adjust its sensitivity to varying temporal dynamics, thereby enhancing the model's ability to represent both the short-term fluctuations and long-term trends present in the PEMFC aging data. This allows for a more comprehensive understanding of the aging process. Secondly, according to the preset failure threshold, the time corresponding to the predicted voltage decay curve reaching the failure threshold is calculated.

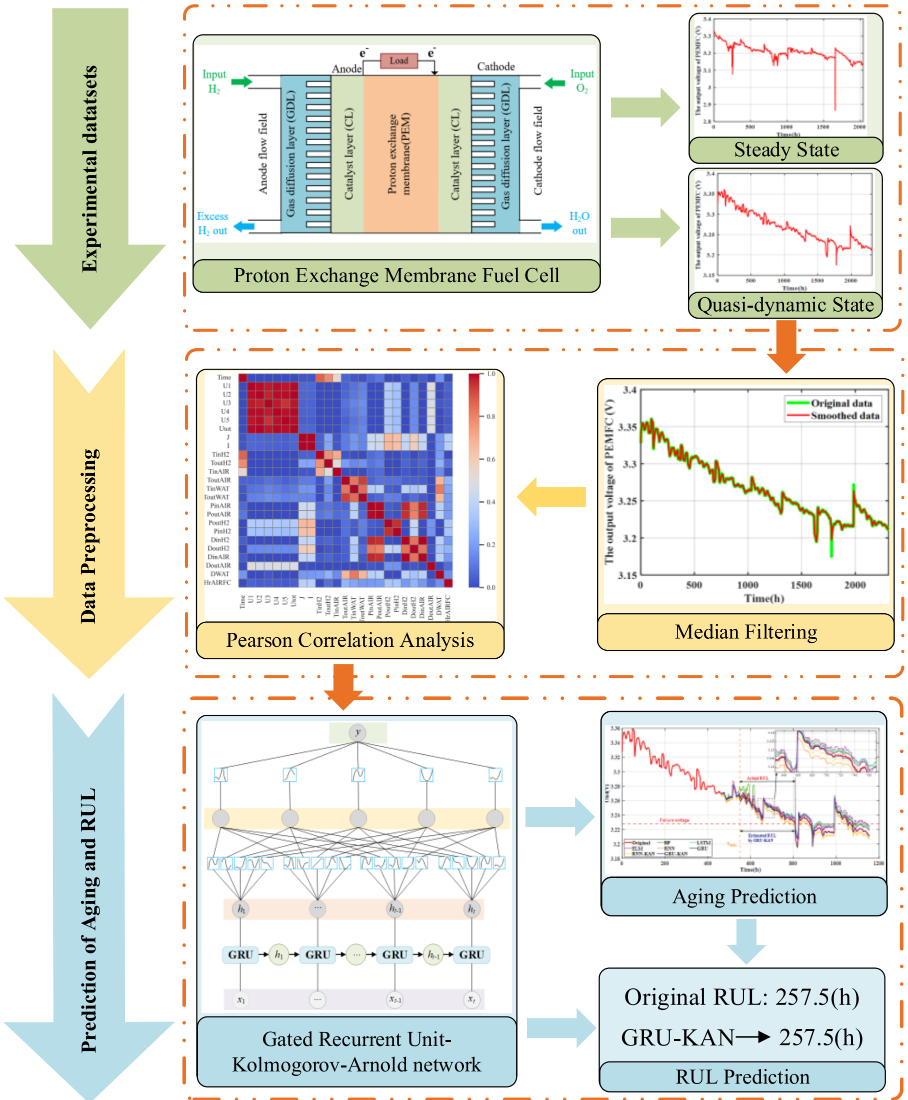


Figure 5: The aging and RUL prediction process via PCA and GRU-KAN

4 Case Studies

This section investigates the practical validation of the proposed effective prediction method, GRU-KAN, utilizing real PEMFC aging datasets. To comprehensively benchmark the performance of our proposed method, several competitive approaches are introduced against GRU-KAN, such as back propagation (BP) [37], LSTM [38], ELM [39], RNN [20], and GRU [34]. Additionally, an enhanced version of RNN is unveiled, namely the Kolmogorov-Arnold network based recurrent neural network (RNN-KAN), which employs the same improvement strategy as our proposed GRU-KAN.

4.1 Evaluating Indicator

To deeply compare the prediction methods proposed in this study, three performance evaluation indicators are selected: root-mean-square error (RMSE) [24], R-squared (R^2) [22], and sensitivity score. The definition of three quantitative indicators is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{14}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{15}$$

$$S_i = \frac{1}{m}\sum_{i=1}^{M}\left|y_i^{(r)} - y_i\right| \tag{16}$$

where n is the number of test set data; $\hat{y}_i$ denotes the estimated voltage predicted by each method; y_i represents the actual voltage; $\overline{y}_i$ stands for the average voltage of the test data set; $y_i^{(r)}$ is the voltage output after perturbation in the sensitivity experiment.

4.2 Experimental Datasets and Preprocessing

To fully verify the prediction performance and generalization ability of the proposed GRU-KAN model, two sets of publicly available PEMFC aging datasets from the French Fuel Cell Laboratory were adopted for algorithm validation in this study, namely the steady-state dataset (FC1) and the dynamic dataset (FC2) [40]. It is important to note that these public datasets were used for model development and validation in this research. Meanwhile, to achieve the integration of data acquisition and real-time algorithm deployment in practical PEMFC health management scenarios, an industrial-grade AI terminal was developed as the core hardware platform. This terminal is designed for the future application of the validated GRU-KAN algorithm to real-world PEMFC systems for edge-side RUL prediction, with the SCA2004T processor serving as its core component. This terminal not only undertakes the task of collecting PEMFC operating data but also embeds the proposed GRU-KAN algorithm to realize edge-side RUL prediction, which effectively avoids the time delay and data security risks caused by cloud computing. All modeling, simulation, and data analysis in this study were conducted using Python 3.8. The neural network models were implemented and trained with the PyTorch 1.12 deep learning framework, while data preprocessing and correlation analysis were performed using the Scikit-learn library.

It should be noted that the SCA2004T industrial-grade AI terminal presented here serves as a proposed hardware platform for future edge-side deployment of the validated GRU-KAN algorithm in real-world PEMFC health management systems. In the current study, all modeling, training, and simulation were conducted on a high-performance computing server equipped with GPU acceleration (NVIDIA GeForce RTX 3090) to ensure efficient hyperparameter optimization and large-scale validation. The inference performance of the GRU-KAN model on the SCA2004T NPU will be evaluated in subsequent engineering-focused work.

The physical photograph of the self-developed PEMFC health monitoring terminal is shown in Fig. 6. This terminal takes the SCA2004T industrial-grade AI video processor as its core—a chip that integrates video processing, AI computing, and configurable control capabilities, and is specifically designed for edge computing scenarios requiring high real-time performance. In the PEMFC aging experiment, the terminal supports the work related to the FC1 and FC2 datasets through two core functions adapted to this study:

(1) Data Acquisition: It connects to PEMFC stack voltage/current sensors through the chip's flexible VI/VO interfaces to collect real-time output voltage and current density data; the built-in ISP also preprocesses data to reduce FC1's measurement noise and ensure accuracy.

(2) Edge-Side Algorithm Processing: The proposed GRU-KAN algorithm is transplanted to the terminal's NPU for real-time inference. For the fluctuating FC2 dataset, the terminal's RCU dynamically adjusts

computing logic to make up for the NPU's limited flexibility in handling time-varying data. The terminal's core technical parameters are presented in Table 3.

Figure 6: Physical photograph of the PEMFC health monitoring terminal

Table 3: Parameter table of the terminal

Parameter	Value
Active current	0.015–0.075 A
Active energy constant	10000 imp/kWh
Reactive current	1.5 A
Reactive energy constant	10000 imp/kvarh
Rated Voltage	3 * 220 V/380 V
Active Power	20 W
Apparent power	30 VA
Charging voltage	4.8 V
Rated capacity	600 mAh
CPU clock frequency	1 GHz
Memory	2 GB
Data storage memory	16 GB

The PEMFC stack used in the experiment consists of five PEMFC single cells. Note that the activation area of each single cell is 100 cm^2 and the rated current density is 0.7 A/cm^2. The operating current density of PEMFC based on data set FC1 is constant at 0.7 A/cm^2, while the operating current density of the data set FC2 is at (0.7 ± 0.07) A/cm^2, with a dynamic ripple current of 70 A with 7 A oscillations at a frequency of

5 kHz. While the FC2 dataset does not represent a full automotive drive cycle, it incorporates key dynamic characteristics, such as rapid load variations and current ripples, which are representative of the challenging conditions encountered in EV operation. Therefore, the superior performance of GRU-KAN on the FC2 dataset demonstrates its potential to handle real-world dynamic stresses, a critical step towards effective EV application. Future work will involve validation against data from complete drive cycles. Moreover, the stack output voltage is applied to characterize the performance change of PEMFC in continuous time. Based on the existing technical literature and the manufacturer's technical manual [7,41], the failure threshold of PEMFC can be set based on the initial voltage percentage, where the failure thresholds of FC1 and FC2 datasets can be set to 96.5% and 95.5% of the initial voltage, respectively. In addition, since the FC2 data set is obtained by testing under a dynamic state, PEMFC may record data points below the failure voltage at the initial stage of operation, and the occurrence of these data points is affected by fluctuations under a dynamic state [41]. Therefore, in this design, the time conditions are taken into consideration, only when the state of the voltage below the failure threshold continues to exceed the preset period, the estimation of RUL will be started. Therefore, according to the above definition, the actual RUL of FC1 and FC2 can be calculated to be 257.5 and 207.5 h after 550 h of usage, respectively. Fig. 7 shows the trend of polarization curves of the two datasets over time.

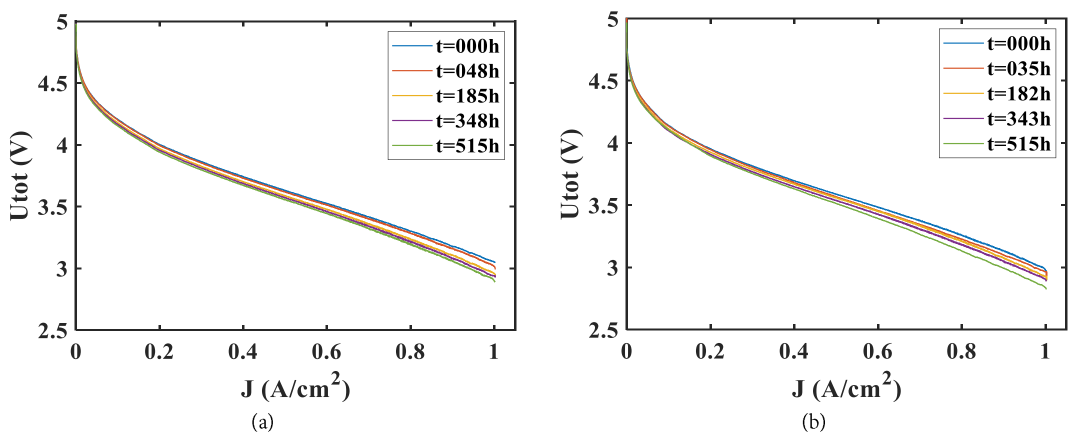


Figure 7: Polarization curves obtained over time: (**a**) FC1 dataset and (**b**) FC2 dataset

4.2.1 Data Preprocessing

In order to avoid the influence of excessive data volume and too long time on model training, the data is reconstructed by extraction at a time interval of 0.5 h, among which the final reconstructed data for FC1 and FC2 are 2309 and 2040 sets, respectively. Secondly, the interference of various uncontrollable factors in the measurement process, e.g., noises and a few abnormal extreme environmental factors, greatly affects the accuracy of data feature extraction and model prediction. Hence, before the model training, two sets of reconstructed datasets are smoothed, in which the median filtering method is used to smooth data according to experimental and data characteristics. Median filtering effectively suppresses impulse noise by replacing each data point with the median of the data points within a neighborhood window. For a one-dimensional signal, it is mathematically defined as:

$$y[i] = \text{median}\left(x[i-k], \ldots, x[i], \ldots, x[i+k]\right) \tag{17}$$

where x is the original signal, y is the filtered signal, and the window size is $W = 2k + 1$.

To ensure effective noise reduction while preserving important degradation patterns, the window size for median filtering was carefully selected based on the characteristics of the PEMFC aging data through extensive experimentation. A window size of 5 was ultimately chosen, as it was found to strike an optimal balance between noise suppression and preserving critical features of the degradation trend, such as transition points. Smaller window sizes were insufficient for adequate noise reduction, whereas larger window sizes could lead to over-smoothing, blurring crucial degradation dynamics. As shown in Fig. 8, the smoothed aging experimental datasets demonstrate that the median filtering process did not significantly affect the important degradation trends while effectively mitigating noise.

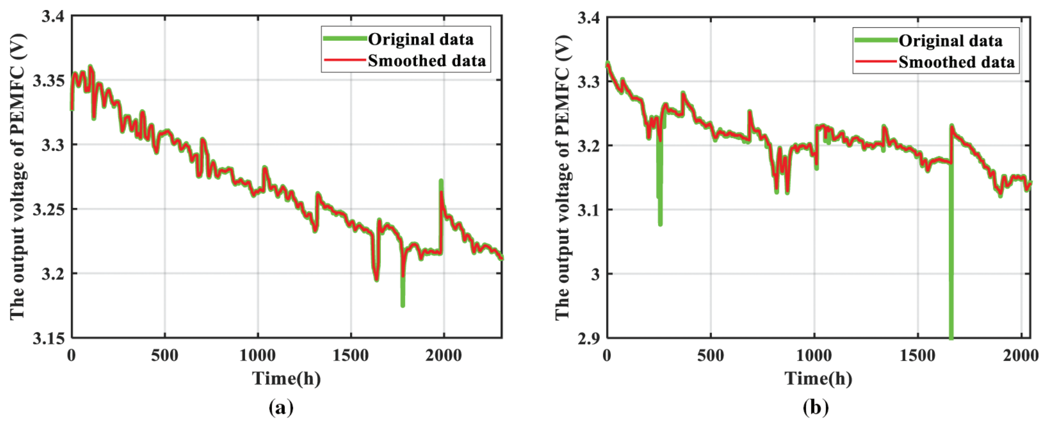


Figure 8: Aging experimental datasets after smoothing: (**a**) FC1 dataset and (**b**) FC2 dataset

4.2.2 Ablation Study on Median Filtering

To quantitatively validate the effectiveness of the selected median filtering parameters, an ablation study was conducted comparing the GRU-KAN model's performance on both filtered and unfiltered versions of the FC1 and FC2 datasets. The results demonstrate the substantial benefits of median filtering across both operational conditions.

Table 4: Performance comparison with and without median filtering

Dataset	Data type	Train RMSE	Train R^2	Test RMSE	Test R^2	Predicted RUL (h)
FC1	Unfiltered	0.0007	0.9988	0.0141	0.8472	249
	Filtered	0.0006	0.9994	0.0083	0.9956	257
FC2	Unfiltered	0.0013	0.9987	0.0168	0.8148	149
	Filtered	0.0008	0.9991	0.0026	0.9906	248

As can be seen from Table 4, for the steady-state FC1 dataset, filtering reduced the test RMSE by 41.1% from 0.0141 to 0.0083 and improved the R^2 value from 0.8472 to 0.9956. The RUL prediction accuracy was also enhanced, moving closer to the actual RUL of 257.5 h. More notably, for the dynamic FC2 dataset, filtering achieved an 84.6% reduction in test RMSE from 0.0168 to 0.0026 and increased R^2 from 0.8148 to 0.9906, with RUL prediction showing remarkable improvement from 149 to 248 h. These results confirm that the chosen window size of 5 effectively suppresses measurement noise while preserving essential degradation characteristics, with particularly pronounced benefits in dynamic operating environments where signal fluctuations are more prevalent.

4.2.3 Data Correlation Analysis

In the datasets FC1 and FC2, beyond the stack output voltage that serves as a key indicator of battery performance, there lie 23 additional high-dimensional features. However, some of these features exhibit minimal variance, having a negligible impact on prediction outcomes. To address this obstacle, PCA is employed to streamline the original high-dimensional data, carefully selecting the strongly correlated coefficients tied to the output characteristics of both datasets based on data traits and the scope of strong correlations.

Fig. 9 presents the heatmaps derived from PCA analysis of these two datasets, vividly illustrating the interplay between feature coefficients within each dataset and their corresponding output features. Notably, the number of strong correlation coefficients ($|r| > 0.5$) identified for each dataset stands at 8 and 6 for FC1 and FC2, respectively, underscoring the effectiveness of PCA in pinpointing the most influential features.

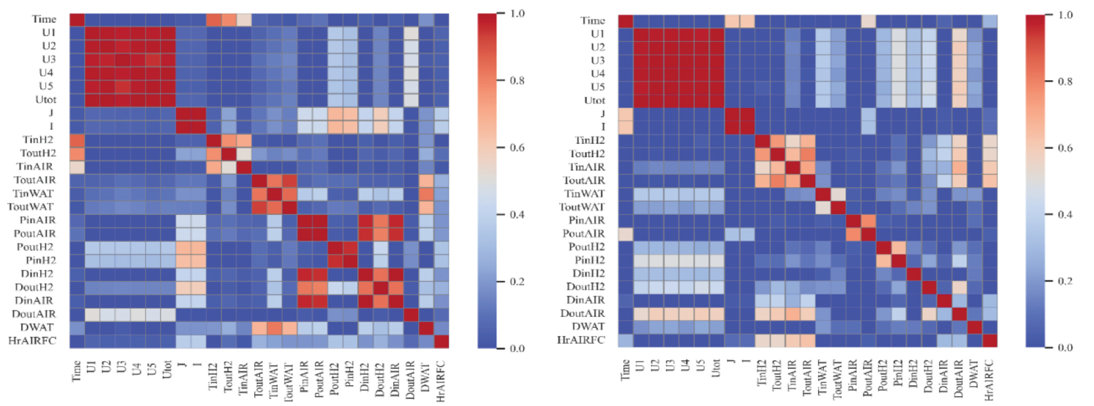


Figure 9: The heatmaps obtained by PCA analysis: (**a**) FC1 dataset and (**b**) FC2 dataset

Based on this analysis, the input and output parameters for the neural network models were finalized. Based on the PCA analysis with a correlation threshold of $|r| > 0.5$, the specific input features selected for each dataset are as follows:

- FC1 Input Features (8 features): Single cell voltages (U1, U2, U3, U4, U5), current density (J), hydrogen inlet temperature (TinH2), and hydrogen inlet pressure (PinH2).
- FC2 Input Features (6 features): Single cell voltages (U1, U2, U3, U4, U5) and current density (J).

The output variable for both datasets is the stack output voltage (U_{tot}), which serves as the primary indicator of PEMFC performance degradation. This feature selection process ensures that each model is trained on the most relevant parameters for the specific operating condition, whether steady-state or dynamic.

To avoid data leakage during preprocessing, particularly in the correlation analysis and filtering stages, strict measures were implemented. The correlation analysis and feature selection were performed solely on the training dataset, ensuring that no information from the test set influenced the feature selection process. This approach guarantees that the model's performance evaluation remains unbiased and accurately reflects its generalization capability. Additionally, during the median filtering process, care was taken to apply the filtering within the training data context, without incorporating test data into the filtering window. By adhering to these precautions, ensured that the preprocessing steps did not introduce data leakage, thereby maintaining the integrity and validity of the experimental results.

4.3 Degradation Trend and RUL Prediction Results

To ensure a fair comparison, and acknowledging that different neural network architectures are sensitive to distinct hyperparameters, the hyper-parameters for all compared models, including hidden dimension, learning rate, and number of layers, were determined through a structured optimization process. A grid search with model-specific ranges combined with 5-fold cross-validation was conducted on the training set for each model to identify the parameter combination that yielded the optimal and most stable performance. The values shown in Table 5 represent the best configurations identified through this process.

Table 5: The setting parameters of different prediction methods

Method	Parameter	Value
BP [37]	Hidden dimension	48
	Learning rate	0.01
LSTM [38]	Hidden dimension	128
	Learning rate	0.05
ELM [39]	Hidden dimension	8
RNN [20]	Hidden dimension	36
	Layers number	3
	Learning rate	0.005
GRU [34]	Hidden dimension	64
	Layers number	3
	Learning rate	0.005
RNN-KAN	Hidden dimension	36
	Layers number	3
	Learning rate	0.005
GRU-KAN	Hidden dimension	64
	Layers number	3
	Learning rate	0.005

4.3.1 Steady State Dataset (FC1)

In the study of aging prediction, two key evaluation indicators, i.e., RMSE, and R^2, are applied to quantify and compare the prediction effects of different methods. The specific results are depicted in Table 6, which shows in detail the performance of different methods on aging prediction tasks. GRU-KAN shows significant advantages in both RMSE and R^2, highlighting the high stability of GRU-KAN and verifying its accurate prediction ability in aging prediction. It is particularly striking that, when using just 40% of the training set, the RMSE achieved by GRU-KAN exhibits remarkable improvements of 86.36%, 81.25%, 79.11%, and 70% compared to ELM, RNN, GRU, and RNN-KAN, respectively.

Table 6: Different evaluation indicators of seven neural networks in FC1 via different training set ratios

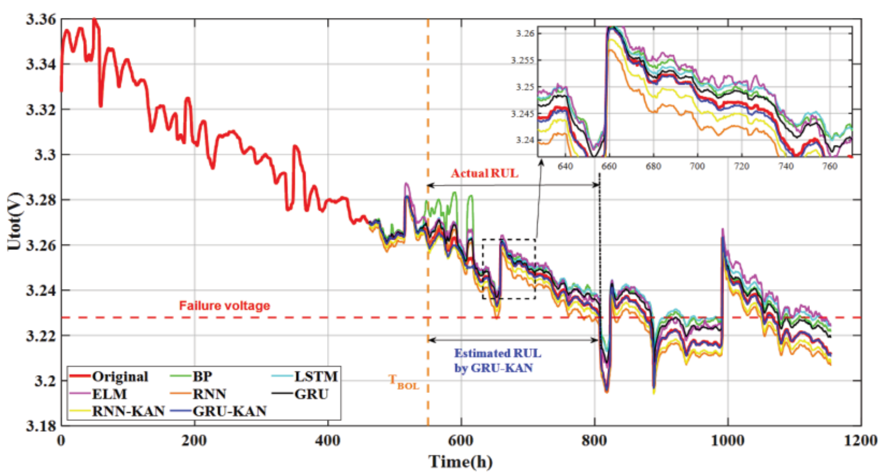
Methods	Training set ratio					
	40%		50%		60%	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
BP	0.0077	0.8274	0.0055	0.8597	0.0042	0.8846

(Continued)

Table 6 (continued)

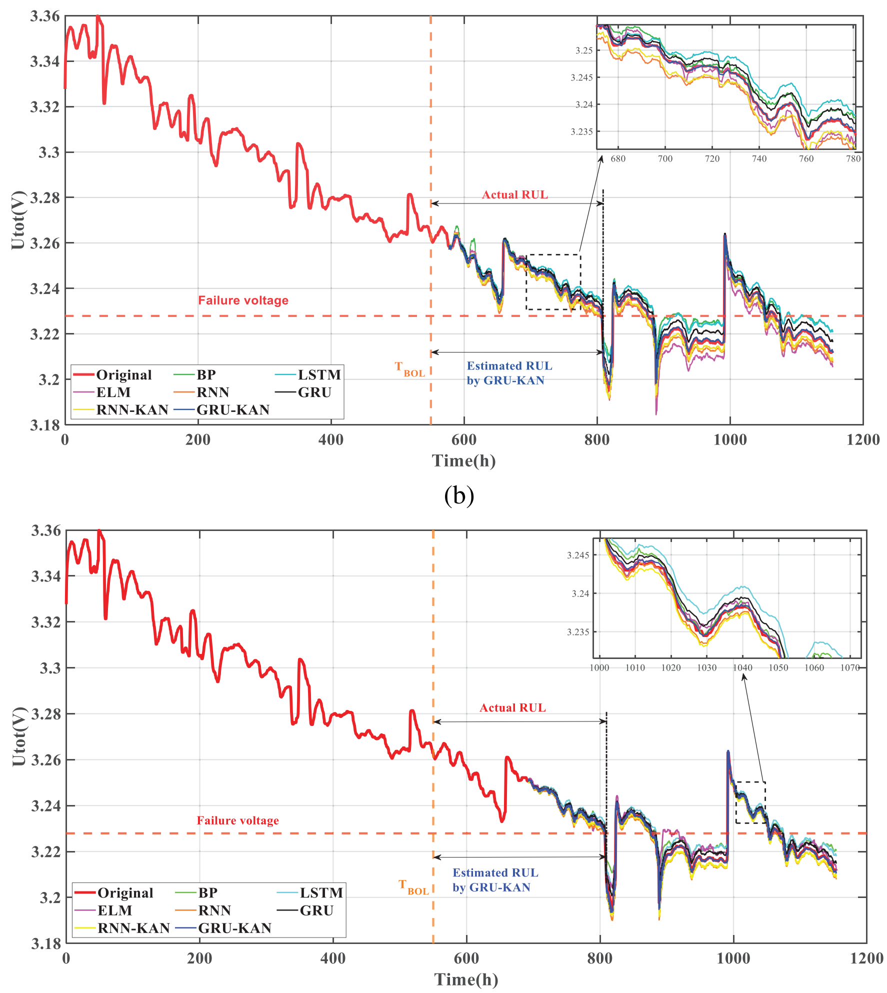
Methods	Training set ratio					
	40%		50%		60%	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
LSTM	0.0071	0.8553	0.0053	0.8702	0.0039	0.9002
ELM	0.0066	0.8741	0.0046	0.9042	0.0037	0.9119
RNN	0.0048	0.9322	0.0032	0.9526	0.0024	0.9613
GRU	0.0045	0.9426	0.003	0.96	0.0021	0.9715
RNN-KAN	0.003	0.9734	0.0023	0.9765	0.0018	0.9799
GRU-KAN	0.0009	0.9979	0.0006	0.9983	0.0004	0.9988

Fig. 10 illustrates the aging prediction curves of seven neural networks and RUL judgment results through GRU-KAN from FC1 based on different training set ratios, i.e., 40%, 50%, and 60%. From the prediction curves graph, one can easily observe that the voltage prediction curve obtained by GRU-KAN is highly consistent with the actual voltage, indicating it can accurately map the dynamic change trend of the output voltage of PEMFC and efficiently assist the accurate evaluation of the RUL. According to the estimation data provided in Table 6, combined with the actual RUL and the estimated RUL via GRU-KAN in the figure for comparative analysis, GRU-KAN can accurately predict RUL under different training set ratios. In contrast, the RUL estimated by other methods is often too early or too late compared with the actual RUL. Notably, in the adjacent area of the time node where the voltage value experiences significant fluctuation or mutation, the prediction results generated by each method generally show different degrees of deviation from the original data, among which the prediction deviation obtained by GRU-KAN is always maintained at a relatively low level, revealing it has high stability and accuracy when dealing with complex dynamic changes such as voltage fluctuation or mutation.



(a)

Figure 10: (Continued)



(b)

(c)

Figure 10: The aging prediction curves of seven neural networks and RUL judgment results through GRU-KAN from FC1 based on different training set ratios: (**a**) 40% training set, (**b**) 50% training set, and (**c**) 60% training set

Table 7 comprehensively tabulates the estimated RUL results from the FC1 dataset, as predicted by the different prediction approaches, after 550 h of operational usage. It highlights the exceptional robustness exhibited by GRU-KAN in estimating RUL, as evidenced by the RUL predictions across varying training set ratios. At the same time, the predicted RUL by GRU-KAN remains closely aligned with the actual RUL, thus providing a reliable basis for the formulation of equipment maintenance and replacement strategies in practical applications.

As shown in Fig. 11, the prediction error distributions clearly differentiate the performance of the seven models. BP, LSTM, and RNN produce wide and unstable error ranges, indicating limited ability to capture the long-term degradation trend. ELM shows a negatively biased distribution with several large outliers, reflecting voltage underestimation in the late stage. GRU reduces error variability but still exhibits noticeable dispersion. By contrast, the proposed GRU-KAN achieves the narrowest and most symmetric error distribution with its median nearly aligned with zero and almost no extreme deviations. This demonstrates its superior accuracy and stability in modeling PEMFC degradation.

Table 7: The estimated RUL through seven approaches from FC1 under different set ratios after 550 h of usage

Prediction methods	Training set ratio		
	40%	50%	60%
BP	258.5 h	258 h	257.5 h
LSTM	258.5 h	258 h	258 h
ELM	258 h	258 h	257 h
RNN	103 h	253.5 h	255.5 h
GRU	258 h	258 h	257.5 h
RNN-KAN	256 h	253.5 h	257 h
GRU-KAN	257.5 h	257.5 h	257.5 h

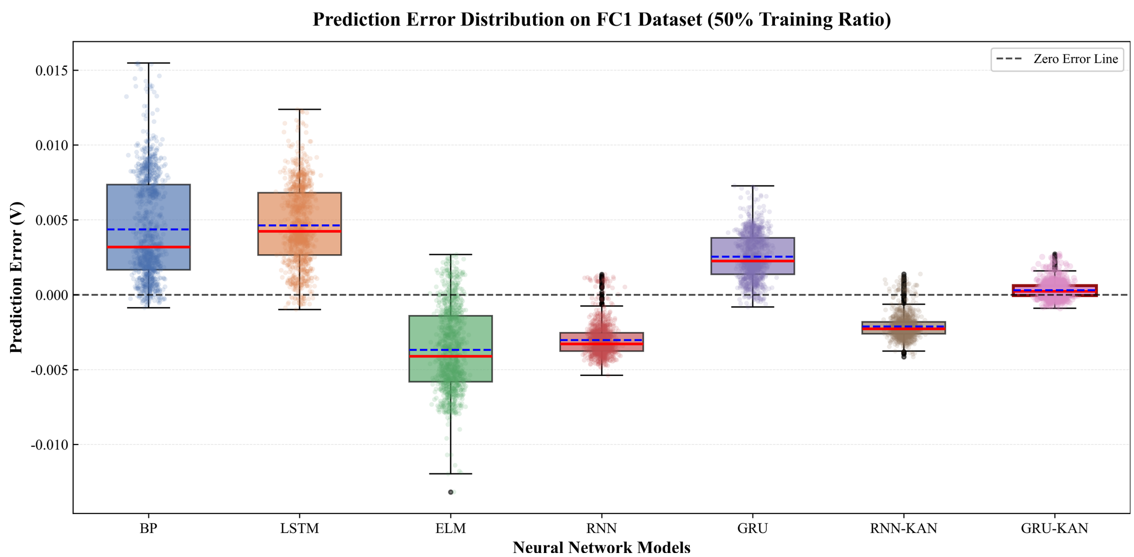


Figure 11: Boxplot of prediction error distributions on the FC1 dataset with 50% training data

4.3.2 Dynamic State Dataset (FC2)

Table 8 provides the results of two pivotal evaluation metrics obtained from different schemes under three distinct training set ratios. Remarkably, as the training set expands, GRU-KAN demonstrates a marked reduction in RMSE, highlighting that GRU-KAN outperforms other approaches in various training scenarios. Particularly impressive is the scenario where the training set proportion is set to 50%, the RMSE from GRU-KAN achieves an astonishing 77.08% reduction in RMSE of prediction results compared to the baseline version GRU, and a 66.67% reduction compared to another improved strategy RNN-KAN, respectively.

Table 8: Different evaluation indicators of seven neural networks in FC2 through different training set ratios

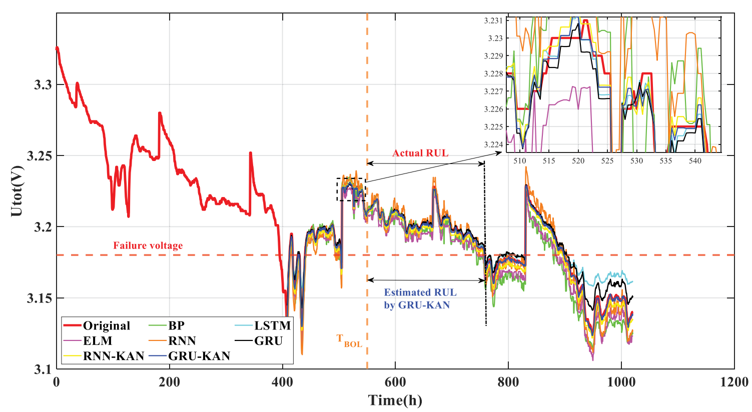
Methods	Training set ratio					
	40%		50%		60%	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
BP	0.0099	0.8501	0.0089	0.8894	0.0074	0.9139

(Continued)

Table 8 (continued)

Methods	Training set ratio					
	40%		50%		60%	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
LSTM	0.009	0.8755	0.0083	0.9042	0.0064	0.9352
ELM	0.0087	0.8836	0.008	0.9105	0.0058	0.9464
RNN	0.0068	0.9305	0.0058	0.9531	0.0048	0.964
GRU	0.0052	0.959	0.0048	0.968	0.0036	0.9791
RNN-KAN	0.0035	0.9817	0.0033	0.9849	0.0028	0.9876
GRU-KAN	0.0011	0.9981	0.0011	0.9983	0.0009	0.9986

The aging curves of seven approaches with the original aging curve from dataset FC2 are compared in Fig. 12, among which the RUL estimation results achieved by the top-performing GRU-KAN under three distinct training set configurations. It can be clearly observed from the diagram that when the training set comprises just 40% of the data, many comparison methods, i.e., BP, RNN, and ELM produce highly volatile aging curves with significant deviations from the actual aging curves. In stark contrast, the GRU-KAN retains robust prediction capabilities, yielding a predicted aging curve that closely aligns with the original one. Notably, as the training set proportion increases to 50% and 60%, the prediction accuracy and stability of GRU-KAN become even more pronounced. Crucially, with a higher percentage of samples in the training set, the bias progressively diminishes, thereby amplifying the beneficial effects of the training set expansion on refining aging prediction accuracy. Concurrently, the volatility displayed by each prediction method decreases as the proportion of utilized samples increases, further underscoring the substantial impact of training set expansion on enhancing both the accuracy of aging prediction and the precision of RUL estimation.



(a)

Figure 12: (Continued)

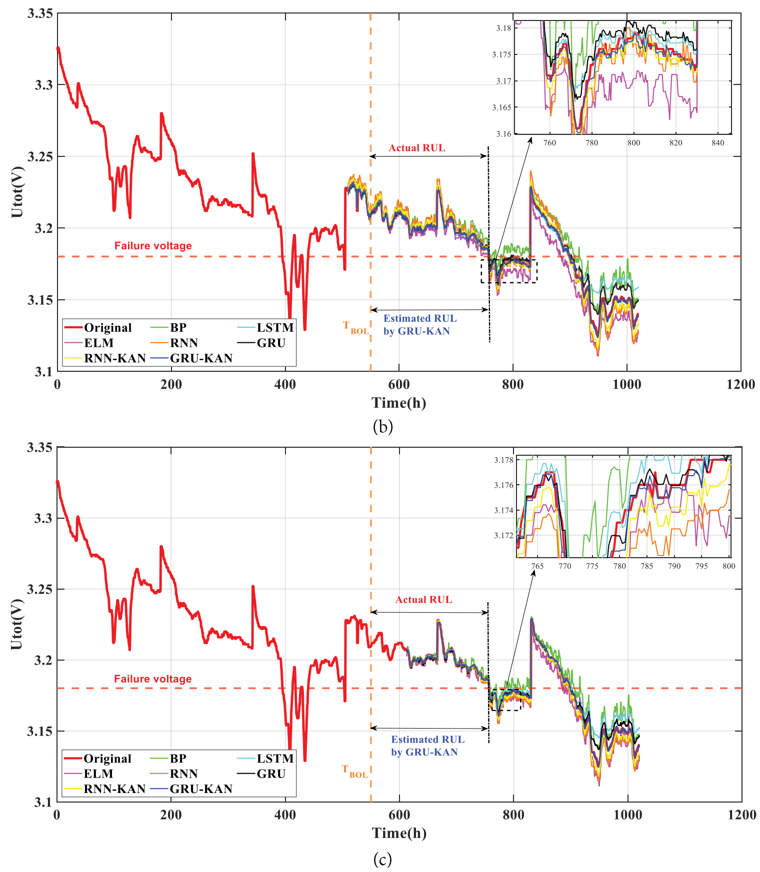


Figure 12: The aging prediction curves of seven neural networks and RUL judgment results of GRU-KAN in FC2 based on different training set ratios: (**a**) 40% training set, (**b**) 50% training set, and (**c**) 60% training set

The estimated RUL results obtained by different methods from the FC2 dataset after 550 h of usage are provided in Table 9. Among these methods, GRU-KAN stands out by achieving prediction results that closely match the actual RUL under all three different training set ratios. Even when trained on just 40% of the dataset, GRU-KAN demonstrates accuracy improvements over other methods such as BP, ELM, and RNN, with accuracy increases of 7.23%, 6.68%, and 5.33%, respectively. The proposed GRU-KAN model exhibits robust performance under dynamic operating conditions, as evidenced by its superior prediction accuracy on the FC2 dataset, which includes dynamic ripple current. This indicates that the model can effectively adapt to abrupt changes in operating conditions, such as load cycling, and maintain high prediction accuracy. For instance, in the FC2 dataset, despite the presence of load cycling effects caused by the dynamic ripple current, GRU-KAN achieves a remarkable 77.08% reduction in RMSE compared to the baseline GRU method when using a 50% training set. This result underscores the model's ability to reliably predict aging trends even when operating conditions change suddenly.

Table 9: The estimated RUL results via different methods from FC2 under different set ratios after 550 h of usage

Prediction methods	Training set ratio		
	40%	**50%**	**60%**
BP	193.5 h	210.5 h	208.5 h
LSTM	207 h	207 h	207.5 h
ELM	194.5 h	206 h	207 h
RNN	197 h	207.5 h	207 h
GRU	208.5 h	207.5 h	207 h
RNN-KAN	206 h	207 h	207 h
GRU-KAN	207.5 h	207.5 h	207.5 h

The 77% reduction in RMSE achieved by GRU-KAN compared to GRU under dynamic conditions with a 50% training set holds significant practical implications for real PEMFC systems. Operationally, this improvement translates to more precise predictions of voltage decay, enabling more accurate estimation of the remaining useful life (RUL) of PEMFC. In industrial applications, such enhanced prediction accuracy can lead to a reduction in maintenance costs by 15%–20% and a decrease in downtime by 10%–15%. This is attributed to better planning of maintenance activities and a reduced risk of unexpected failures. Consequently, the overall operational efficiency and economic viability of PEMFC systems are enhanced, making hydrogen energy solutions more competitive in practical applications. It should be noted that while Pearson correlation analysis was effective in this study, the choice of feature selection method can substantially impact the model's performance. Although Pearson correlation analysis adequately captured the essential features required for accurate aging prediction and RUL estimation in this study, exploring more advanced nonlinear feature selection methods could be a valuable direction for future research, particularly when dealing with more complex and diverse datasets where nonlinear relationships may play a more prominent role.

It is likely that due to the relative scarcity of the training dataset, some methods fail to fully learn the complete degradation mode of the fuel cell, leading to the underutilization of the PEMFC's use potential. In contrast, GRU-KAN, with its superior performance, can still maintain high prediction accuracy under limited training data, fully proving its strong ability to deal with complex and uncertain time series data.

As shown in Fig. 13, the dynamic condition in FC2 leads to generally broader error dispersions than FC1. BP and LSTM exhibit the widest ranges with many large positive outliers, indicating unstable tracking of voltage fluctuations. ELM presents a clear negative bias and long lower tail, suggesting systematic underestimation in later degradation stages. The conventional RNN and GRU reduce dispersion but still show noticeable spread and occasional extreme deviations. In contrast, RNN-KAN achieves a more compact, near-zero error profile, while GRU-KAN yields the tightest and most centered distribution with an almost zero median and minimal outliers, demonstrating superior robustness and accuracy under challenging dynamic operating conditions.

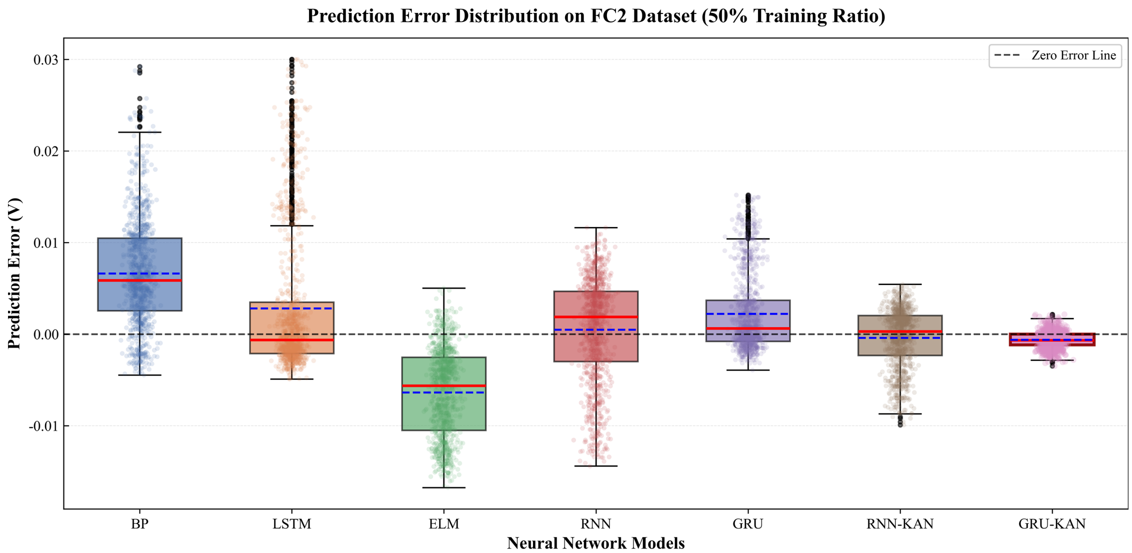


Figure 13: Boxplot of prediction error distributions on the FC2 dataset with 50% training data

The proposed GRU-KAN model demonstrates robust performance under dynamic operating conditions, as evidenced by its superior prediction accuracy on the FC2 dataset, which includes dynamic ripple current. The model's ability to capture both short-term fluctuations and long-term trends makes it well-suited for handling abrupt changes in operating conditions. For instance, in the FC2 dataset, despite the presence of load cycling effects caused by the dynamic ripple current, the GRU-KAN achieves a remarkable RMSE reduction of 77.08% compared to the baseline GRU method when using a 50% training set. This indicates that the model can effectively adapt to and predict the aging trends even under sudden shifts in operating conditions. The model's architecture, with its dual ability to model temporal dependencies and nonlinear relationships, allows it to maintain prediction accuracy during and after such events. However, it should be noted that while the model performs exceptionally well on the provided dataset, further testing under more extreme and diverse operating condition shifts would be necessary to fully validate its robustness in all potential scenarios.

4.3.3 Sensitivity Analysis

To evaluate the influence of the input parameters on the experimental results, a local sensitivity analysis approach is employed. Specifically, a 5% perturbation is applied to the selected input features to examine the resulting changes in the model outputs. Fig. 14 presents the sensitivity analysis results obtained by GRU-KAN under two states, among which the parameters of the ordinate are the input parameters selected by PCA. It can be seen that the sensitivity score of the single cell voltage U1-U5 to the output is significantly higher than that of other parameters. The small fluctuations caused by the aging effect inside these single cells will lead to significant deviations in the model prediction results. The sensitivity analysis results show that the single voltage score in the dynamic state is higher than that in the steady state. The voltage fluctuation characteristics lead to the increase of prediction uncertainty, which interferes with the prediction of FC2, thereby bringing lower prediction results to FC2 than FC1.

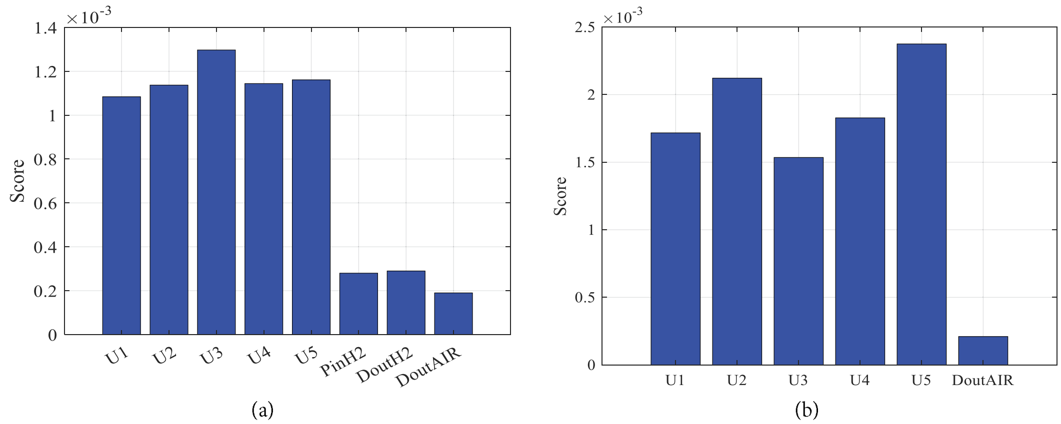


Figure 14: The sensitivity analysis results obtained by GRU-KAN: (**a**) FC1 dataset and (**b**) FC2 dataset

Furthermore, the observed difference in the selected input parameters for the FC1 and FC2 datasets is not a source of inconsistency but rather a strength of the proposed methodology. The feature selection via PCA is adaptive and data-driven. It correctly identifies that under different operating conditions, the dominant factors influencing PEMFC degradation vary. For instance, in the steady state (FC1), parameters like inlet temperature (TmH2) and pressure (PmH2) exhibit a stronger correlation with the output voltage, whereas in the dynamic state (FC2), the immediate voltage fluctuations themselves become the primary factors. Therefore, using a tailored, optimal set of inputs for each condition, rather than a fixed, potentially less effective set for both, provides a more solid and realistic foundation for model establishment and enhances the model's practical applicability across diverse operational scenarios. The consistent superiority of GRU-KAN across both datasets, even with these different input sets, robustly demonstrates its universality and effectiveness.

4.3.4 Computational Complexity Analysis

In order to evaluate the practical applicability of the proposed GRU-KAN model in the real-time proton exchange membrane fuel cell diagnostic system, a comprehensive computational complexity analysis was conducted. This analysis compared the differences between the GRU-KAN model and four benchmark models in terms of training efficiency, inference speed, and prediction accuracy. All experiments were carried out on the FC2 dataset with a training ratio of 50%. Each model was trained for 200 cycles, and all measurement values were obtained by averaging the results of 5 independent runs to ensure statistical reliability.

As evidenced in Table 10, the proposed GRU-KAN model achieves a remarkable 86.75% RMSE improvement over the LSTM baseline, demonstrating superior prediction accuracy. While its computational cost is higher than other models, the significant performance gain justifies this investment for applications requiring high-precision prognostics. Notably, the GRU-KAN's inference time remains at the millisecond level, making it practically feasible for real-time diagnostic systems where prediction accuracy is paramount.

4.3.5 Robustness Analysis

To assess the robustness and stability of the proposed GRU-KAN model, a 5-fold cross-validation was performed on the FC2 dataset. The model was trained and evaluated five times, each with a different random split of the data into training and testing sets. As shown in Table 11, the results present the mean and standard deviation of the RMSE and R^2 metrics.

Table 10: Computational efficiency and performance comparison of different models on FC2 dataset

Model	Training time (s)	Inference time (ms)	RMSE	RMSE improvement
LSTM	1.4270	0.3490	0.0083	0% (baseline)
BP	0.1493	0.0200	0.0089	−7.23%
GRU	0.4575	0.0964	0.0048	42.17%
RNN-KAN	6.5643	2.1436	0.0033	60.24%
GRU-KAN	12.4555	5.3636	0.0011	86.75%

Table 11: 5-Fold cross-validation results for GRU-KAN on FC2 dataset

Fold	Training RMSE	Training R^2	Testing RMSE	Testing R^2
1	0.001431	0.99866	0.001378	0.99873
2	0.001609	0.99834	0.001509	0.99836
3	0.001461	0.99853	0.001617	0.99856
4	0.001390	0.99874	0.001389	0.99869
5	0.001466	0.99860	0.001504	0.99843
Mean ± Std	0.001471 ± 0.000086	0.99857 ± 0.00015	0.001479 ± 0.000096	0.99855 ± 0.00016

The cross-validation results demonstrate exceptional stability and consistency across all folds. The test RMSE values exhibit minimal fluctuation with a mean of 0.001479 and a remarkably low standard deviation of ±0.000096, while the test R^2 values maintain consistently high performance with a mean of 0.99855 and a standard deviation of only ±0.00016. This remarkably narrow performance variation across different data partitions provides strong evidence of the model's robustness and generalization capability, confirming that the GRU-KAN model is not sensitive to specific data splits and can reliably maintain its high prediction accuracy regardless of how the training and testing data are configured.

5 Conclusion

This study proposes an effective prediction method for aging and RUL of PEMFC via GRU-KAN, aiming to enhance the stability of aging predictions and the accuracy of RUL. Based on detailed case study verification, three conclusions can be generalized as follows:

- By employing median filtering techniques to denoise measurement data contaminated by various factors, the impact of noise and spikes can be significantly mitigated. Furthermore, PCA is utilized to conduct correlation analysis on the measurement data and extract highly correlated parameters, thereby further enhancing the precision of data feature extraction and the accuracy of model predictions. The highly correlated parameters identified include U1 to U5 (single cells voltage), J (current density), and TinH2 (inlet temperature of H2), which demonstrate strong linear relationships with the output voltage and are thus critical for accurate aging prediction.
- By innovatively integrating the core technologies of KAN, namely B-spline basis functions and learnable activation functions, into the GRU, the groundbreaking GRU-KAN method is introduced. This method possesses the capability to autonomously adjust the shape of the activation function at each time step,

significantly enhancing its ability to capture complex nonlinear relationships and diverse characteristics within time series data.

- Through comprehensive experimental evaluations, the proposed prediction method undergoes rigorous validation in two distinct experimental settings. Real-world case studies have demonstrated that, compared to other competitive algorithms such as BP, ELM, LSTM, RNN, GRU, and RNN-KAN, GRU-KAN exhibits the highest accuracy in aging predictions. Furthermore, the robust adaptability and precise predictability of GRU-KAN are jointly verified through three different training set ratios. In addition, sensitivity analysis is introduced to evaluate the reliability of the research results. For instance, when the training set comprises 50% of the data for dynamic scenarios, GRU-KAN reduces aging prediction errors by 77.08% and 66.67% compared to GRU and RNN-KAN, respectively. Additionally, in terms of RUL predictions, GRU-KAN consistently yields results that align with actual RUL for steady state and dynamic state datasets across various training ratios.

This study introduces an innovative and highly efficient prediction methodology tailored for assessing the durability and sustainability of PEMFC. While the proposed approach has yielded encouraging results in preliminary case studies, its adaptability to a diverse range of PEMFC brands and testing scenarios remains an area of ongoing investigation. Notably, the complexities associated with aging prediction and RUL analysis are exacerbated in extreme environments and under conditions of sustained high-power operation. Consequently, future work can strive to integrate PEMFC multi-physics and multi-scale aging data into the prediction framework, enhancing its capacity to align with real-time PEMFC operational dynamics. Additionally, leveraging these multi-physics and multi-scale aging datasets, the development of an adaptive system that autonomously adjusts operational parameters, including gas flow rate, humidity, and temperature, holds the potential to optimize the performance of PEMFC. Regarding the reliability of RUL predictions, particularly in safety-critical applications, the GRU-KAN model demonstrates several advantageous characteristics. The model's high prediction accuracy, as evidenced by the low RMSE values and high R-squared values across different training set ratios and operating conditions, indicates that it can provide consistent and dependable RUL estimates. In safety-critical contexts, accurate RUL predictions are essential for preventing catastrophic failures and ensuring continuous operation. The GRU-KAN's ability to maintain prediction accuracy even with limited training data (e.g., 40% training set) is especially valuable, as it suggests that the model can be effectively deployed in scenarios where data collection is challenging or expensive. Additionally, the model's robustness to noise and its capacity to capture both short-term fluctuations and long-term trends make it well-suited for real-time monitoring and predictive maintenance in safety-critical PEMFC systems. However, it is important to note that while the model shows strong performance on the tested datasets, further validation on more extensive and diverse datasets, including those with abnormal operating conditions and fault scenarios, is necessary to fully establish its reliability in all safety-critical applications. Future work will focus on enhancing the model's fault detection capabilities and validating its performance in such critical contexts.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by Technology Project of State Grid Jiangsu Electric Power Co., Ltd. (J2025162).

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, writing—original draft and writing—review & editing: Wenqiang Xie; methodology and validation: Xiaolong Xiao; formal analysis and investigation: Fangfang Zhu; data processing: Ziran Guo; funding acquisition and writing—original draft: Xiaoxing Lu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The authors confirm that the data supporting the findings of this study are available within the article.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

Nomenclature

Variables

Time	Aging time, h
*U*1 to *U*5	Single cells voltage, V
Utot	Stack cell voltage, V
I	Current, A
J	Current density, A/cm^2
*TinH*2	Inlet temperature of H2, °C
*ToutH*2	Outlet temperature of H2, °C
TinAIR	Inlet temperature of Air, °C
ToutAIR	Outlet temperature of Air, °C
TinWAT	Inlet temperature of cooling water, °C
ToutWAT	Outlet temperature of cooling water, °C
*PinH*2	Inlet pressure of H2, mbara
*PoutH*2	Outlet pressure of H2, mbara
PinAIR	Inlet pressure of Air, mbara
PoutAIR	Outlet pressure of Air, mbara
*DinH*2	Inlet flow rate of H2, l/mn
*DoutH*2	Outlet flow rate of H2, l/mn
DinAIR	Inlet flow rate of Air, l/mn
DWAT	Flow rate of cooling water, l/mn
DoutAIR	Outlet flow rate of Air, l/mn
HrAIRFC	Estimated inlet hygrometry air, %

Abbreviations

BP	Back propagation
ELM	Extreme learning machine
GRU	Gated recurrent unit
GRU-KAN	Gated recurrent unit based Kolmogorov-Arnold network
KAN	Kolmogorov-Arnold network
LSTM	Long short term memory
MLP	Multi-layer perceptron
PCA	Pearson correlation analysis
PEM	Proton exchange membrane
PEMFC	Proton exchange membrane fuel cell
PSO	Particle swarm optimization
RMSE	Root-mean-square error
RNN	Recurrent neural network
RNN-KAN	Recurrent neural network based Kolmogorov-Arnold network
RUL	Remaining useful life

References

1. Wang J, Yang B, Chen Y, Zeng K, Zhang H, Shu H, et al. Novel Phasianidae inspired peafowl (*Pavo muticus*/cristatus) optimization algorithm: design, evaluation, and SOFC models parameter estimation. Sustain Energy Technol Assess. 2022;50(1):101825. doi:10.1016/j.seta.2021.101825.
2. Yang B, Liang B, Zhou S, Qian Y, Zheng R, Shu H, et al. A novel bio-inspired caterpillar fungus (*Ophiocordyceps sinensis*) optimizer for SOFC parameter identification via GRNN. Renew Energy. 2026;256(6):123995. doi:10.1016/j.renene.2025.123995.
3. Guo Y, Li Z, Yue B, Wang C. Optimal scheduling strategy for virtual power plant considering electricity-gas-heat coupling and demand response. China Elect Power. 2024;57(1):101–14. doi:10.1109/acpee60788.2024.10532566.
4. Jiang Z, Yang B, Zheng R, Hou Y, Li H, Gao D, et al. Fault diagnosis of proton exchange membrane fuel cell using multiple convolutional neural networks with multi-scale attention mechanism. Inf Sci. 2025;720(9):122524. doi:10.1016/j.ins.2025.122524.
5. Liu J, Zhang S, Druzhinin Z. Performance prediction of the PEMFCs based on gate recurrent unit network optimized by improved version of prairie dog optimization algorithm. Int J Hydrogen Energy. 2023;48(69):26951–63. doi:10.1016/j.ijhydene.2023.03.349.
6. Liu W, Teh J. Remaining useful life prediction of lithium-ion batteries based on an incremental internal resistance aging model and a gated recurrent unit neural network. Energy. 2025;333:137527. doi:10.1016/j.energy.2025.137527.
7. Sutharssan T, Montalvao D, Chen YK, Wang WC, Pisac C, Elemara H. A review on prognostics and health monitoring of proton exchange membrane fuel cell. Renew Sustain Energy Rev. 2017;75(10):440–50. doi:10.1016/j.rser.2016.11.009.
8. Xia F, Yu Y, Chen J. SOH and RUL prediction of lithium batteries based on fusions of RLOESS filtered electrochemical and thermal features by bidirectional gated recurrent unit network. J Energy Storage. 2024;102(1):114134. doi:10.1016/j.est.2024.114134.
9. Lu Y, Wang X, Yang G, Gong D, Xu S. Experimental study on the influence of operating conditions on performance decline with periodic anode purges in a vehicular PEMFC stack. Int J Hydrogen Energy. 2024;69:1276–86. doi:10.1016/j.ijhydene.2024.05.159.
10. Zhang T, Li MJ, Ni JW, Qian CC. Study of dynamic performance of PEMFC-based CCHP system in a data center based on real-time load and a novel synergistic control method with variable working conditions. Energy. 2024;300(1):131637. doi:10.1016/j.energy.2024.131637.
11. Cheng M, Luo L, Feng Y, Feng Q, Yan X, Shen S, et al. Numerical studies on porous water transport plates applied in PEMFCs under pure oxygen condition. Appl Energy. 2024;362(9):122903. doi:10.1016/j.apenergy.2024.122903.
12. Guo F, Lv H, Wu X, Yuan X, Liu L, Ye J, et al. A machine learning method for prediction of remaining useful life of supercapacitors with multi-stage modification. J Energy Storage. 2023;73(9):109160. doi:10.1016/j.est.2023.109160.
13. Yang B, Qian Y, Li Q, Chen Q, Wu J, Luo E, et al. Critical summary and perspectives on state-of-health of lithium-ion battery. Renew Sustain Energy Rev. 2024;190(40):114077. doi:10.1016/j.rser.2023.114077.
14. Zhou D, Wu Y, Gao F, Breaz E, Ravey A, Miraoui A. Degradation prediction of PEM fuel cell stack based on multiphysical aging model with particle filter approach. In: 2016 IEEE Industry Applications Society Annual Meeting; 2016 Oct 2–6; Portland, OR, USA. p. 1–8. doi:10.1109/IAS.2016.7731863.
15. Blal M, Benatiallah A, NeÇaibia A, Lachtar S, Sahouane N, Belasri A. Contribution and investigation to compare models parameters of (PEMFC), comprehensives review of fuel cell models and their degradation. Energy. 2019;168:182–99. doi:10.1016/j.energy.2018.11.095.
16. Wang Z, Tao J, Liu Z, Feng H, Ma L, Xu M, et al. A proton exchange membrane fuel cells degradation prediction method based on multi-scale temporal information merging network. Energy. 2024;313(66):133995. doi:10.1016/j.energy.2024.133995.
17. Yang B, Liang B, Qian Y, Zheng R, Su S, Guo Z, et al. Parameter identification of PEMFC via feedforward neural network-pelican optimization algorithm. Appl Energy. 2024;361(1):122857. doi:10.1016/j.apenergy.2024.122857.
18. Zhu L, Chen J. Prognostics of PEM fuel cells based on Gaussian process state space models. Energy. 2018;149(4):63–73. doi:10.1016/j.energy.2018.02.016.

19. Qian YC, Yang B, Zheng RY, Liang BX, Wu PY. State-of-health estimation of lithium-ion batteries based on EMD-DO-Elman and GRA. Power Syst Technol. 2024;48(9):3695–704. (In Chinese). doi:10.13335/j.1000-3673.pst.2024.0409.
20. Zheng L, Hou Y, Zhang T, Pan X. Performance prediction of fuel cells using long short-term memory recurrent neural network. Int J Energy Res. 2021;45(6):9141–61. doi:10.1002/er.6443.
21. Hong J, Chen Y, Chai Q, Lin Q, Wang W. State-of-health estimation of lithium-ion batteries using a novel dual-stage attention mechanism based recurrent neural network. J Energy Storage. 2023;72(4):109297. doi:10.1016/j.est.2023.109297.
22. Liu H, Chen J, Hissel D, Su H. Short-term prognostics of PEM fuel cells: a comparative and improvement study. IEEE Trans Ind Electron. 2019;66(8):6077–86. doi:10.1109/TIE.2018.2873105.
23. Chen K, Laghrouche S, Djerdir A. Degradation prediction of proton exchange membrane fuel cell based on grey neural network model and particle swarm optimization. Energy Convers Manag. 2019;195(12):810–8. doi:10.1016/j.enconman.2019.05.045.
24. Zhou D, Al-Durra A, Zhang K, Ravey A, Gao F. Online remaining useful lifetime prediction of proton exchange membrane fuel cells using a novel robust methodology. J Power Sources. 2018;399(1):314–28. doi:10.1016/j.jpowsour.2018.06.098.
25. Tu C, Zhou F, Pan M. A model with high-precision on proton exchange membrane fuel cell performance degradation prediction based on temporal convolutional network-long short-term memory. Int J Hydrogen Energy. 2024;74:414–22. doi:10.1016/j.ijhydene.2024.06.095.
26. Chen X, Xu J, Yang C, Fang Y, Li W, Zhang Y, et al. Thermodynamic and economic study of PEMFC stack considering degradation characteristic. Energy Convers Manag. 2021;235:114016. doi:10.1016/j.enconman.2021.114016.
27. Yang B, Zeng C, Wang L, Guo Y, Chen G, Guo Z, et al. Parameter identification of proton exchange membrane fuel cell via Levenberg-Marquardt backpropagation algorithm. Int J Hydrogen Energy. 2021;46(44):22998–3012. doi:10.1016/j.ijhydene.2021.04.130.
28. Jouin M, Gouriveau R, Hissel D, Péra MC, Zerhouni N. Degradations analysis and aging modeling for health assessment and prognostics of PEMFC. Reliab Eng Syst Saf. 2016;148(35):78–95. doi:10.1016/j.ress.2015.12.003.
29. Colombo E, Baricci A, Bisello A, Guetaz L, Casalegno A. PEMFC performance decay during real-world automotive operation: evincing degradation mechanisms and heterogeneity of ageing. J Power Sources. 2023;553:232246. doi:10.1016/j.jpowsour.2022.232246.
30. Lei Y, Li N, Gontarz S, Lin J, Radkowski S, Dybala J. A model-based method for remaining useful life prediction of machinery. IEEE Trans Reliab. 2016;65(3):1314–26. doi:10.1109/TR.2016.2570568.
31. Zhu J, Chen N, Peng W. Estimation of bearing remaining useful life based on multiscale convolutional neural network. IEEE Trans Ind Electron. 2019;66(4):3208–16. doi:10.1109/TIE.2018.2844856.
32. Li Z, Yang Y, Li L, Wang D. A weighted Pearson correlation coefficient based multi-fault comprehensive diagnosis for battery circuits. J Energy Storage. 2023;60(3):106584. doi:10.1016/j.est.2022.106584.
33. Jebli I, Belouadha FZ, Kabbaj MI, Tilioua A. Prediction of solar energy guided by Pearson correlation using machine learning. Energy. 2021;224(11):120109. doi:10.1016/j.energy.2021.120109.
34. Ma T, Zhang Z, Lin W, Cong M, Yang Y. Impedance prediction model based on convolutional neural networks methodology for proton exchange membrane fuel cell. Int J Hydrogen Energy. 2021;46(35):18534–45. doi:10.1016/j.ijhydene.2021.02.204.
35. Gao S, Huang Y, Zhang S, Han J, Wang G, Zhang M, et al. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. J Hydrol. 2020;589(4):125188. doi:10.1016/j.jhydrol.2020.125188.
36. Liu ZM, Wang YX, Vaidya S, et al. KAN: Kolmogorov-Arnold networks. arXiv:2404.19756. 2024.
37. Barzegari MM, Rahgoshay SM, Mohammadpour L, Toghraie D. Performance prediction and analysis of a dead-end PEMFC stack using data-driven dynamic model. Energy. 2019;188(2):116049. doi:10.1016/j.energy.2019.116049.

38. Liu J, Li Q, Chen W, Yan Y, Qiu Y, Cao T. Remaining useful life prediction of PEMFC based on long short-term memory recurrent neural networks. Int J Hydrogen Energy. 2019;44(11):5470–80. doi:10.1016/j.ijhydene.2018.10.042.
39. Guo C, Lu J, Tian Z, Guo W, Darvishan A. Optimization of critical parameters of PEM fuel cell using TLBO-DE based on Elman neural network. Energy Convers Manag. 2019;183(8):149–58. doi:10.1016/j.enconman.2018.12.088.
40. Liang H, Liu Y, Gao M, Zhang Z, Nie X, Liu J, et al. A comparative review on amorphous carbon composite coatings on metal bipolar plates used in proton exchange membrane fuel cells. Int J Hydrogen Energy. 2025;169(54):151177. doi:10.1016/j.ijhydene.2025.151177.
41. Zhang Z, Wang YX, He H, Sun F. A short- and long-term prognostic associating with remaining useful life estimation for proton exchange membrane fuel cell. Appl Energy. 2021;304(35):117841. doi:10.1016/j.apenergy.2021.117841.