

Article

Interpretable Transformer-Based Voltage Degradation Prediction of Proton Exchange Membrane Fuel Cells Under Constant-Current Operation

Fengyan Yi ¹, Xing Shu ¹, Jinming Zhang ², Zongjing Huang ², Junling Zhang ², Hongtao Gong ^{1,*},
Xiangya Liu ², Shuaihua Wang ¹ and Jiaming Zhou ^{2,*}

¹ School of Automotive Engineering, Shandong Jiaotong University, Jinan 250357, China; yify@sdjtu.edu.cn (F.Y.)

² School of Mechanical and Electrical Engineering, Weifang University of Science and Technology, Weifang 262700, China

* Correspondence: ght010208@163.com (H.G.); zhoujiaming@wfust.edu.cn (J.Z.)

Abstract

Accurate voltage degradation prediction is essential for health management and lifetime extension of proton exchange membrane fuel cell (PEMFC) systems. During long-term constant-current operation, stack voltage evolves nonlinearly and is influenced by coupled variations in temperature, pressure, flow rate, and humidity, while many deep learning-based models lack physical interpretability. This study proposes an interpretable Transformer-based framework for PEMFC voltage degradation prediction under constant-current operation. The framework integrates outlier correction, interpolation, Savitzky–Golay filtering, Z-score normalization, sliding-window reconstruction, Transformer-based prediction, and feature-ablation interpretation. Using multivariate sensor measurements and historical voltage as inputs, the Transformer was compared with RNN, LSTM, and GRU baselines under identical preprocessing and evaluation conditions. The models were evaluated chronologically by continuously applying the sliding-window model over the held-out final 20% of the aging sequence. The Transformer achieved the best performance, with MAE of 8.2×10^{-4} , RMSE of 1.18×10^{-3} , MAPE of 0.0256%, and R^2 of 0.9961. Compared with the second-best RNN model, it reduced MAE, RMSE, and MAPE by 9.89%, 7.81%, and 9.86%, respectively. Feature ablation showed that flow- and pressure-related variables contributed 50.08% and 26.78% of the total importance, respectively.

Keywords: PEMFC; degradation prediction; Transformer; feature ablation; interpretable machine learning



Academic Editor: Fabio Corti

Received: 1 June 2026

Revised: 13 July 2026

Accepted: 27 July 2026

Published: 28 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Proton exchange membrane fuel cells (PEMFCs) are considered a promising power source for hydrogen fuel cell vehicles because of their high efficiency, low operating temperature, fast response, and zero-emission characteristics [1–3]. However, durability limitations still hinder the practical deployment of PEMFCs in large-scale commercial applications [4]. During long-term operation, irreversible degradation of components such as the catalyst layer, proton exchange membrane, gas diffusion layer, and bipolar plates gradually causes voltage decay and performance loss [5–7]. Therefore, accurate degradation prediction is essential for health management, maintenance planning, and lifetime extension of PEMFC systems.

PEMFC degradation is a nonlinear and multi-factor coupled process involving electrochemical reactions, reactant transport, water balance, and thermal management [8,9]. Under constant-current operation, the external load and environmental variables are relatively stable, which helps reduce transient disturbances and makes the intrinsic voltage degradation trend easier to observe. Nevertheless, steady-state degradation still evolves slowly over long time scales and is affected by multiple operating variables, such as gas flow rate, pressure, temperature, and humidity [10]. These variables reflect the reactant supply, product removal, mass transport state, and water–thermal regulation of the stack. Consequently, it remains challenging to build a prediction model that can accurately capture degradation-related voltage evolution from historical operating data while maintaining physical interpretability [11].

Existing PEMFC degradation prediction methods can generally be divided into model-based, data-driven, and hybrid approaches [8,12–14]. Model-based methods predict degradation by establishing mathematical models that describe internal aging processes, including mechanism-based and empirical models. Mechanism-based methods usually describe degradation from electrochemical and physicochemical perspectives. For example, Singh et al. [15] developed a physical model for chemical membrane degradation and evaluated the influence of radical attack on PEMFC performance. Ao et al. [16] proposed a frequency-domain Kalman filtering method combined with a voltage degradation model, achieving efficient PEMFC prognosis under different operating conditions. Empirical models, by contrast, usually fit degradation trends from experimental data. Pei et al. [17] proposed a rapid evaluation method for automotive fuel cell lifetime by considering dynamic load, start–stop cycles, and idling conditions. Wang et al. [9] combined semi-empirical modeling and data-driven correction to improve the stability of long-term degradation prediction. More recently, Meng et al. [18] proposed a polarization loss decomposition-based online health state estimation method for PEMFCs, in which degradation-related indicators such as electrochemical surface area and equivalent resistance were estimated to provide direct electrochemical insight into fuel cell aging. Although model-based methods usually have clear physical meanings, their practical application is still limited by parameter identification difficulty, model uncertainty, high computational cost, and weak adaptability to complex operating environments [19,20].

Data-driven methods provide an alternative way to predict PEMFC degradation without requiring detailed knowledge of internal physicochemical mechanisms. By learning degradation patterns directly from measured signals such as voltage, current, impedance, pressure, and flow rate, machine learning and deep learning models have shown strong potential in nonlinear prognostic tasks. Tang et al. [10] proposed an online degradation model based on transfer learning to improve prediction performance under changing data distributions. Sun et al. [11] developed a CNN–LSTM hybrid framework in which CNN was used to extract spatial features from multivariate sensor data, and LSTM was used for temporal sequence modeling. He et al. [21] proposed a temporal causal inference graph neural network framework for probabilistic remaining useful life estimation of PEMFCs, improving prediction accuracy, robustness, and uncertainty quantification. Jin et al. [22] developed a long-term PEMFC degradation prediction method based on an improved echo state network with adaptive fuzzy sampling, which enhanced the extraction of key voltage degradation information. These studies demonstrate the effectiveness of data-driven methods for PEMFC lifetime prediction. However, recurrent architectures such as RNN, LSTM, and GRU process sequence information sequentially through hidden-state propagation, which limits parallel computation and may affect their ability to model flexible interactions among multivariate observations.

Hybrid prediction methods have also attracted increasing attention because they combine physical knowledge with data-driven learning to improve robustness and generalization [23,24]. Chen et al. [25] proposed a least squares support vector machine–random process flow framework to integrate the advantages of model-driven and data-driven methods. Liu et al. [26] developed a hybrid algorithm based on particle swarm optimization and adaptive neuro-fuzzy inference for PEMFC degradation trend prediction and remaining useful life estimation. Ke et al. [27] combined degradation mechanisms with deep neural networks to predict PEMFC lifetime degradation under dynamic load cycles. Du et al. [28] constructed a hybrid prediction model combining voltage and mechanism models, in which dynamic weights were assigned according to local prediction errors. Meng et al. [29] considered the recovery phenomenon of reversible voltage loss in fuel cell life prediction, highlighting the need to distinguish recoverable voltage decay from irreversible aging in PEMFC prognostics. Although hybrid strategies improve prediction reliability by incorporating physical constraints or mechanism knowledge, their performance still depends on the completeness and accuracy of prior degradation models.

For PEMFC aging data under constant-current operation, voltage degradation evolves gradually and is influenced by historical voltage and multiple sensor-measured operating variables [9]. Therefore, an effective prediction model should be able to capture multivariate temporal interactions and identify the most informative observations within the input window. The Transformer provides a suitable architecture because its self-attention mechanism directly models relationships among different time steps and adaptively evaluates their relevance without recursive hidden-state propagation [30].

Recent studies have applied Transformer-based architectures to PEMFC degradation prediction, such as temporal fusion Transformer with covariates and Transformer-based prediction considering reversible voltage recovery [31,32]. However, these studies mainly focus on prediction accuracy or specific degradation phenomena, while the feature-time decision mechanism of Transformer models still requires further investigation.

Another important issue is the limited interpretability of existing deep learning-based PEMFC prognostic models. Most studies focus mainly on statistical prediction metrics, such as MAE, RMSE, MAPE, and R^2 , while paying insufficient attention to why the model produces a specific prediction and which operating variables dominate the predicted degradation trend [33]. For safety-critical applications such as fuel cell vehicles, prediction accuracy alone is insufficient. A reliable prognostic model should also provide physically understandable evidence that its decision logic is consistent with PEMFC degradation mechanisms. Feature attribution methods offer a feasible way to analyze model behavior [34]. In particular, feature ablation can directly evaluate the change in predicted voltage after replacing or removing a specific input feature. By extending ablation analysis from the feature dimension to the temporal dimension, it becomes possible to identify not only key operating parameters but also influential historical time positions.

Motivated by these considerations, this study proposes an interpretable Transformer-based framework for online PEMFC voltage degradation prediction under constant-current operation. The proposed framework first conducts systematic data preprocessing and sliding-window reconstruction to obtain high-quality multivariate degradation sequences. A Transformer prediction model is then developed to capture temporal dependencies from operating parameters and historical voltage information. To evaluate its effectiveness, the proposed model is compared with representative recurrent neural network baselines, including RNN, LSTM, and GRU. More importantly, a feature ablation-based interpretation strategy is introduced to analyze the model's decision mechanism from three complementary perspectives: global feature contribution, temporal position sensitivity, and feature–time joint attribution. In this way, the proposed framework is evaluated not only

in terms of prediction accuracy but also in terms of the physical plausibility of its learned degradation-related patterns.

- An interpretable Transformer-based model is developed for PEMFC voltage degradation prediction over a chronologically held-out late-life interval. Under identical sliding-window and evaluation settings, it achieves better prediction performance than the representative RNN, LSTM, and GRU baselines.
- A feature ablation-based interpretation framework is proposed to reveal the model decision mechanism from feature, time, and feature-time dimensions. The attribution results identify flow-related and pressure-related variables as dominant factors in constant-current degradation prediction, which is consistent with PEMFC mass transport degradation characteristics.

The remainder of this paper is organized as follows. Section 2 introduces the experimental dataset and data preprocessing procedures. Section 3 describes the Transformer-based prediction model and the feature ablation interpretation method. Section 4 presents the prediction performance comparison and interpretability results. Section 5 discusses the physical implications of the attribution results and summarizes the main conclusions.

2. Experimental Dataset and Preprocessing

2.1. PEMFC Lifetime Degradation Dataset

The experimental data used in this study were obtained from the public IEEE PHM 2014 fuel cell degradation dataset provided by the FCLAB research center [35]. The dataset contains aging data for two fuel cell stacks. Because the present study focuses on degradation prediction during constant-current operation, only FC1 was used for model development and evaluation. The main parameters of FC1 experiment are shown in Table 1. FC1 operated at a constant stack current of 70 A for approximately 1154 h. Although the current setpoint remained constant, the sensor-measured temperature, pressure, flow rate and humidity variables exhibited temporal variation and were retained as multivariate model inputs. The condition is therefore referred to as a constant-current condition. The stack consisted of five cells, each with an active area of 100 cm²; the nominal operating temperature, hydrogen pressure, and relative humidity were 54 degrees C, 1.3 bar, and 50%, respectively. Periodic characterization tests temporarily interrupted the normal aging condition and could produce abrupt voltage changes and local disturbances in the raw sequence.

Table 1. Main operating conditions of the FC1 degradation experiment.

Operating Condition	Constant-Current Condition (FC1)
Number of single cells	5
Active area	100 cm ²
Rated stack current	70 A
Operating temperature	54 °C
Hydrogen pressure	1.3 bar
Relative humidity	50%

The stack output voltage was selected as the health indicator and prediction target because it directly reflects the electrochemical performance degradation of the fuel cell. As the stack ages, irreversible changes in the mass transport pathways, gas diffusion layer, catalyst layer, and membrane electrode assembly gradually lead to voltage decay. Therefore, accurate prediction of the stack voltage trend can provide useful information for degradation assessment and health management.

2.2. Feature Selection

The original dataset contains multiple monitored variables, including stack-level electrical signals, single-cell voltages, gas supply variables, pressure variables, temperature variables, flow variables, and humidity-related parameters. However, not all measured variables are suitable as model inputs for stack-level degradation prediction. The individual cell voltages U_1-U_5 were excluded because the stack voltage U was selected as the prediction target, and direct use of single-cell voltages could introduce redundant information highly correlated with the target variable. The current density J was also removed because it is directly derived from the current and active area. Since both hydrogen and air are compressible fluids, the relationship between their pressure and flow rate cannot be directly described by Bernoulli's equation, nor can they be substituted for each other. Therefore, both pressure and flow rate were filtered out and treated as characteristic parameters to be analyzed. Ultimately, 18 characteristic parameters were retained, as shown in Table 2.

Table 2. Selected variables for constant-current PEMFC degradation prediction.

Classification	Feature Parameters	Symbol	Unit
Voltage	Fuel cell output	U	V
	Current	I	A
Temperature	Hydrogen inlet	T_{inH2}	°C
	Hydrogen outlet	T_{outH2}	
	Air inlet	T_{inAIR}	
	Air outlet	T_{outAIR}	
	Coolant inlet	T_{inWAT}	
	Coolant outlet	T_{outWAT}	
Pressure	Hydrogen inlet	P_{inH2}	Mbar
	Hydrogen outlet	P_{outH2}	
	Air inlet	P_{inAIR}	
	Air outlet	P_{outAIR}	
Flow rate	Hydrogen inlet	D_{inH2}	L/min
	Hydrogen outlet	D_{outH2}	
	Air inlet	D_{inAIR}	
	Air outlet	D_{outAIR}	
	Coolant	D_{inWAT}	

2.3. Data Preprocessing and Sliding Window Construction

The raw fuel cell aging experiment data is inevitably affected by factors such as sensor noise, acquisition errors, instantaneous abnormal disturbances, and switching between characterization tests during long-term acquisition. To improve the quality of input data and ensure the stability of model training, this paper sequentially performs outlier processing, filtering and denoising, data standardization, and sliding window sample construction.

An outlier detection method based on interquartile range (ICR) is employed in a voltage series. Let the first and third ICRs of the voltage series be Q_1 and Q_3 , respectively. Then, the ICR is defined as:

$$IQR = Q_3 - Q_1 \quad (1)$$

A voltage value at a given moment is considered an outlier if it satisfies the following formula:

$$U_i < Q_1 - 1.5IQR \text{ or } U_i > Q_3 + 1.5IQR \quad (2)$$

For detected outliers, a linear interpolation method is used for replacement to avoid unreasonable impacts of outliers on model training and prediction results. This method

does not rely on specific data distribution assumptions and is applicable to the asymmetric distribution and local fluctuations that may exist in long-term fuel cell aging data.

To reduce the interference of high-frequency measurement noise on the extraction of lifetime decay trends, this paper employs the Savitzky–Golay (SG) filtering method to smooth the voltage sequence. SG filtering, through polynomial least-squares fitting within a local sliding window, effectively suppresses random noise while preserving the local slope variations and trend characteristics of the voltage decay curve. Compared to simple moving average filtering, SG filtering is less prone to over-smoothing of degradation inflection points, making it more suitable for preprocessing PEMFC voltage decay trends. This paper sets the filter window length to 101 and the local polynomial order to 3. The voltages before and after filtering are shown in Figure 1.

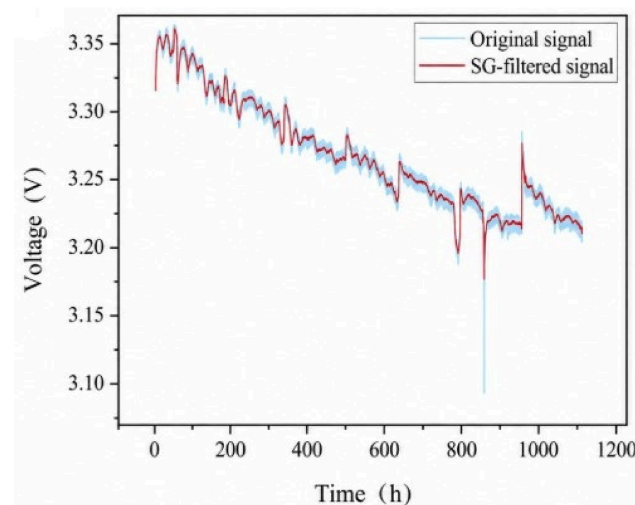


Figure 1. SG filtering noise reduction effect.

Because the dimensions and numerical ranges of different input variables vary significantly—for example, the orders of magnitude of current, pressure, temperature, and flow rate are not the same—directly inputting them into the model can easily lead to variables with larger numerical ranges dominating the training process. To eliminate the impact of these dimensional differences on model training, this paper uses the Z-score normalization method to normalize each input variable.

$$x_{norm} = \frac{x - \mu}{\sigma} \quad (3)$$

In the formula, x represents the original variable value, and μ and σ are the mean and standard deviation of the variable on the training set, respectively. It is important to note that all standardized parameters are calculated only based on the training set; the test set is transformed using the same mean and standard deviation to avoid data leakage. The predicted target voltage is also independently standardized, and after model output, it is restored to the actual voltage value through inverse standardization.

A sliding window method is used to convert continuous time series into supervised learning samples. Let the sliding window length be L and the input feature dimension be F , then the input of the t -th sample can be represented as:

$$X_t = [x_{t-L}, x_{t-L+1}, \dots, x_{t-1}] \in \mathbb{R}^{L \times F} \quad (4)$$

The corresponding prediction target is the current stack voltage:

$$y_t = U_t \quad (5)$$

The prediction task is defined as chronological sliding-window voltage prediction. For each sample, the L observations within the current historical window of the multivariate operating variables and voltage are used to estimate the corresponding target stack-voltage value. The window length was fixed at $L = 50$ and the step size at 1 for all compared models. The original high-frequency data were downsampled at fixed intervals to reduce redundancy. The first 80% of the sequence was used for training, and the final 20% was retained as a chronologically unseen late-life test interval. This partitioning method more realistically simulates actual online prediction scenarios, i.e., using historical data to predict future degradation states, rather than randomly partitioning the data and disrupting the chronological order.

3. Methods

3.1. Transformer-Based Degradation Prediction Model

Under constant-current operating conditions, PEMFC voltage degradation evolves slowly and continuously and is jointly affected by multiple operating variables. To capture the temporal dependencies among voltage, flow rate, pressure, temperature, and humidity, this study develops a task-adapted Transformer model for long-term degradation-trajectory prediction. Rather than directly adopting the original encoder–decoder Transformer, an encoder-only architecture is employed because the objective is to continuously track voltage degradation from fixed-length historical windows, instead of performing sequence-to-sequence generation. The model consists of a multivariate input projection layer, learnable positional encoding, stacked Transformer encoder layers, and a regression output layer.

Given input window $X_t \in \mathbb{R}^{L \times F}$, the original input features are projected into a high-dimensional latent space through a linear mapping:

$$H_0 = X_t W_e + b_e \quad (6)$$

where W_e and b_e represent the weight matrix and bias term of the input embedding layer, respectively. Since the Transformer itself does not contain a recursive structure and cannot directly perceive sequence position information, positional encoding is required. This paper employs learnable positional encoding, which is added to the input embedding:

$$Z_0 = H_0 + P \quad (7)$$

where P is the learnable location encoding matrix.

The core of the Transformer encoder is a multi-head self-attention mechanism [36]. For the input representation Z , the query matrix, key matrix, and value matrix are first obtained through a linear transformation:

$$Q = ZW^Q, K = ZW^K, V = ZW^V \quad (8)$$

The scaled dot product attention calculation method is as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (9)$$

Here, d_k is the dimension of the key vector. The attention mechanism dynamically allocates weights according to relationships among positions in the historical window, enabling the model to select the temporal information most relevant to the subsequent voltage estimate.

To enhance the model's ability to represent features from different subspaces, this paper employs a multi-head attention mechanism:

$$head_i = \text{Attention}(ZW_i^Q, ZW_i^K, ZW_i^V), i = 1, \dots, h \quad (10)$$

$$\text{MultiHead}(Z) = \text{Concat}(head_1, \dots, head_h)W^O \quad (11)$$

where $head_i$ denotes the i -th attention head; h is the number of attention heads; W_i^Q , W_i^K , and W_i^V are the head-specific query, key, and value projection matrices, respectively; $\text{Concat}(\cdot)$ denotes the concatenation operation; and W^O is the output projection matrix.

$$Z' = \text{LayerNorm}(Z + \text{MultiHead}(Z)) \quad (12)$$

$$Z_{out} = \text{LayerNorm}(Z' + \text{FFN}(Z')) \quad (13)$$

The feedforward sublayer applies two successive linear transformations with a nonlinear activation function inserted between them:

$$\text{FFN}(Z') = \text{ReLU}(Z'W_1 + b_1)W_2 + b_2 \quad (14)$$

After passing through multiple Transformer encoders, the hidden state of the last time step is taken as the timing representation of the current window:

$$h_t = Z_{out}^{(L)} \quad (15)$$

Final predicted output voltage value through the fully connected layer:

$$\hat{y}_t = W_o h_t + b_o \quad (16)$$

The Transformer model used in this paper mainly consists of an input linear projection layer, a learnable positional encoder, two Transformer encoder layers, and an output fully connected layer. Input features are first mapped to a 64-dimensional latent space. The encoder uses four attention heads, the feedforward network has a hidden dimension of 256, and Dropout is introduced to reduce the risk of overfitting. The model output is a standardized voltage prediction, which is then recovered to the actual voltage value through inverse standardization.

To provide a clearer and more reproducible description of the task-adapted Transformer, its detailed architecture, key parameters, and dimensional transformations are summarized in Table 3.

Table 3. Detailed architecture and dimensional transformations of the proposed Transformer-based regression model.

Module	Layer Type	Key Parameters	Output Dimensions
linear projection layer	Linear	$17 \rightarrow 64$	(B, 50, 64)
positional encoding	Learnable parameter	$L_{max} = 500,$ $d_{model} = 64$	(B, 50, 64)
Transformer encoder $\times 2$	Multi-Head Attention + FFN	$n = 4, d_{ff} = 256,$ $dropout = 0.2$	(B, 50, 64)
Take the last time step	Last-time-step indexing	(H[:, -1, :])	(B, 64)
Fully connected layer	Linear	$64 \rightarrow 1$	(B, 1)

3.2. Baseline Models and Implementation Settings

To evaluate the effectiveness of the proposed Transformer model, three representative recurrent neural network models, including RNN, LSTM, and GRU, were selected as baseline models. These models are commonly used in time-series prediction tasks and have also been widely adopted in degradation modeling and health prognostics. The RNN model was used as a basic recurrent architecture, while LSTM and GRU were introduced as gated recurrent models to improve the ability to capture temporal dependencies.

For a fair comparison, all baseline models used the same input features, window length, chronological train-test split, and evaluation metrics as the proposed Transformer. Each baseline contained two recurrent layers, 128 hidden units, and a dropout rate of 0.2. The representation at the final input position was passed to a fully connected layer to generate the corresponding stack-voltage estimate. This sliding-window procedure was repeated across the entire held-out late-life interval, and all outputs were inverse-normalized before the evaluation metrics were calculated.

The proposed Transformer model consisted of two encoder layers, four attention heads, a hidden dimension of 64, a feed-forward dimension of 256, and a dropout rate of 0.2. The same prediction task and preprocessing procedure were applied to all compared models to ensure that the performance differences mainly resulted from model architecture rather than data processing or experimental settings.

3.3. Evaluation Metrics

To comprehensively evaluate the voltage degradation prediction performance, four commonly used regression metrics were adopted: root mean squared error (RMSE), coefficient of determination (R^2), mean absolute error (MAE), and mean absolute percentage error (MAPE). These metrics evaluate the prediction performance from different perspectives. RMSE reflects the overall magnitude of prediction errors, MAE measures the average absolute deviation, MAPE evaluates the relative prediction error, and R^2 indicates the goodness of fit between the predicted and measured voltage values.

The MAE is calculated as

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (17)$$

The RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (18)$$

The MAPE is calculated as

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (19)$$

where N is the number of test samples, y_i is the measured stack voltage, and $\hat{y}_i$ is the predicted stack voltage.

The coefficient of determination R^2 is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (20)$$

where $\bar{y}$ is the mean value of the measured voltage. Smaller MAE, RMSE, and MAPE values indicate better prediction accuracy, while a larger R^2 indicates stronger fitting ability and better explanation of the voltage degradation trend.

3.4. Feature Ablation-Based Interpretability Analysis

Although the Transformer model can provide accurate voltage degradation prediction, its internal decision process is not directly transparent. To improve model interpretability and examine whether the learned prediction logic is physically meaningful, a feature ablation-based attribution method was introduced. The basic idea of feature ablation is to replace a specific input feature at a specific historical time step with a reference value and then measure the change in the model output. A larger output change indicates that the corresponding feature-time position contributes more strongly to the prediction.

For an original input sample X , the prediction of the trained Transformer model is expressed as

$$\hat{y} = f(X) \quad (21)$$

where $f(\cdot)$ denotes the trained prediction model. When the j -th feature at the t -th time step is replaced by its reference value, the perturbed input is denoted as $X'_{j,t}$. The attribution value of this feature-time position is calculated as

$$I_{j,t} = |f(X) - f(X'_{j,t})| \quad (22)$$

This setting removes the sample-specific information of the selected feature while keeping the perturbed input within a reasonable data distribution. Since the current remained constant throughout the FC1 constant-current aging experiment, it contained little variation and was excluded from the interpretability analysis. Therefore, feature ablation analysis was performed on the remaining 16 operating variables.

To identify the dominant operating variables in degradation prediction under constant-current operation, the global importance of each feature was obtained by averaging the ablation effect over all test samples and all historical time steps:

$$I_j = \frac{1}{NL} \sum_{n=1}^N \sum_{t=1}^L |f(X^{(n)}) - f(X'_{j,t})| \quad (23)$$

where N is the number of test samples and L is the input sequence length. The normalized contribution of the j -th feature was further calculated as

$$C_j = \frac{I_j}{\sum_{j=1}^{F'} I_j} \times 100\% \quad (24)$$

where $F' = 16$ denotes the number of operating variables used for attribution analysis. This global attribution result provides a quantitative ranking of the input variables and helps determine whether the model mainly relies on flow, pressure, temperature, or humidity-related information for voltage degradation prediction.

In addition to feature-level attribution, temporal position sensitivity was analyzed to determine which historical time steps contributed most to the model prediction. The importance of the t -th time step was calculated by averaging the ablation effect over all analyzed features and all test samples:

$$I_t = \frac{1}{NF'} \sum_{n=1}^N \sum_{j=1}^{F'} |f(X^{(n)}) - f(X'_{j,t})| \quad (25)$$

The normalized temporal contribution was defined as

$$C_t = \frac{I_t}{\sum_{t=1}^L I_t} \times 100\% \quad (26)$$

This analysis reveals whether the Transformer model mainly depends on recent operating states or earlier historical information. Here, each historical time step denotes a relative temporal position within the input window and does not refer to a forecasting horizon. For PEMFC degradation under constant-current operation, such temporal attribution is useful for understanding how the model uses historical information to infer the current degradation trend.

Furthermore, to jointly examine feature importance and temporal sensitivity, a feature-time attribution matrix was constructed as

$$A = [I_{j,t}] \in \mathbb{R}^{F' \times L} \quad (27)$$

Each element $I_{j,t}$ represents the attribution value of the j -th feature at the t -th historical time step. By visualizing this matrix as a heatmap, the key feature-time combinations used by the model can be identified. This joint attribution analysis provides a more fine-grained explanation than global feature ranking or temporal importance analysis alone. It can reveal whether the model prediction is based on all input variables across the whole time window or mainly depends on a small number of key variables at specific recent time points.

Through the above feature ablation procedure, the Transformer model is interpreted from three complementary perspectives: feature contribution, temporal sensitivity, and feature-time joint attribution. This enables the prediction model to be evaluated not only in terms of numerical accuracy, but also in terms of the physical plausibility of its learned decision mechanism under constant-current PEMFC aging conditions.

4. Results and Discussion

4.1. Prediction Performance Comparison

To evaluate prediction performance during the held-out late-life portion of the FC1 constant-current aging sequence, the Transformer was compared with RNN, LSTM, and GRU models. All models used the same preprocessing procedure, input-window length, chronological train-test split, and evaluation metrics to ensure a consistent comparison.

Figure 2 shows the voltage prediction results of four models under constant-current conditions of FC1. Overall, the FC1 voltage exhibits a relatively slow decreasing trend with increasing operating time, with slight fluctuations in local areas caused by characterization tests or system disturbances. The RNN model can track the overall voltage decay trend well, but still shows some deviation in local fluctuation regions. While the LSTM model can capture some time dependencies, its predicted curve deviates significantly from the true value in some periods, indicating its limited ability to characterize the long-term constant-current degradation process. The GRU model's prediction error fluctuates relatively large, and its response to local voltage changes within the test interval is not stable enough. In contrast, the Transformer model's predicted curve matches the true voltage curve best, accurately tracking the long-term voltage degradation trend and maintaining low prediction deviation in local fluctuation regions.

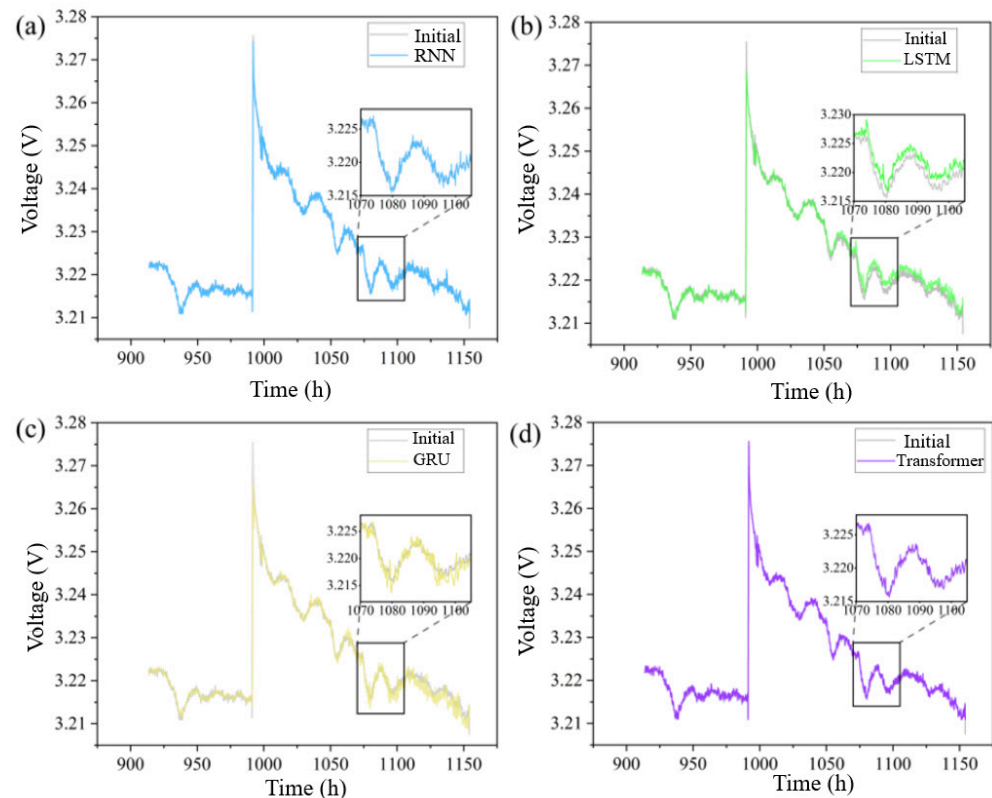


Figure 2. Comparison of FC1 prediction results from four methods: (a) RNN, (b) LSTM, (c) GRU, (d) Transformer.

As can be further seen from the evaluation metrics in Table 4, the Transformer achieved the best prediction performance under the FC1 constant-current condition, with MAE, RMSE, and MAPE of 8.2×10^{-4} , 1.18×10^{-3} , and 0.0256%, respectively, and a coefficient of determination R^2 of 0.9961. Compared with the second-best performing RNN, the Transformer reduced MAE by 9.89%, RMSE by 7.81%, and MAPE by 9.86%, indicating that it not only has higher fitting accuracy but also more stable prediction error. The prediction performance of LSTM and GRU is significantly lower than that of Transformer. The FC1 comparison results in the original paper show that Transformer outperforms RNN, LSTM, and GRU on all evaluation metrics, achieving a prediction accuracy of $R^2 = 0.9961$.

Table 4. Four methods for evaluating FC1 metrics.

Method	MAE	RMSE	MAPE	R^2
RNN	9.1×10^{-4}	1.28×10^{-3}	0.0284%	0.9943
LSTM	1.31×10^{-3}	2.11×10^{-3}	0.0407%	0.9786
GRU	2.36×10^{-3}	2.76×10^{-3}	0.0733%	0.9658
Transformer	8.2×10^{-4}	1.18×10^{-3}	0.0256%	0.9961

4.2. Global Feature Contribution Analysis

From the perspective of global feature contributions, the impact of different operating parameters on voltage prediction varies significantly. The air outlet flow rate (D_{outAIR}) is the most important, contributing approximately 24.8%; followed by the hydrogen outlet flow rate (D_{outH2}), contributing approximately 15.3%. In terms of parameter categories, flow-related features have the highest cumulative contribution, reaching 50.1%; pressure-related features contribute 26.8%; temperature-related features contribute 18.7%; and humidity-related features contribute the least, only 4.4%. This result indicates that under constant

current, the model mainly relies on flow and pressure information reflecting gas transport and discharge status to determine voltage degradation trends.

From the category-level attribution results in Figure 3, flow-related variables account for the largest proportion, reaching 50.1% of the total contribution. Pressure-related variables rank second, with a contribution of 26.8%, followed by temperature-related variables at 18.7% and humidity-related variables at 4.4%. This distribution suggests that the Transformer model mainly uses gas transport-related information to infer the voltage degradation trend.

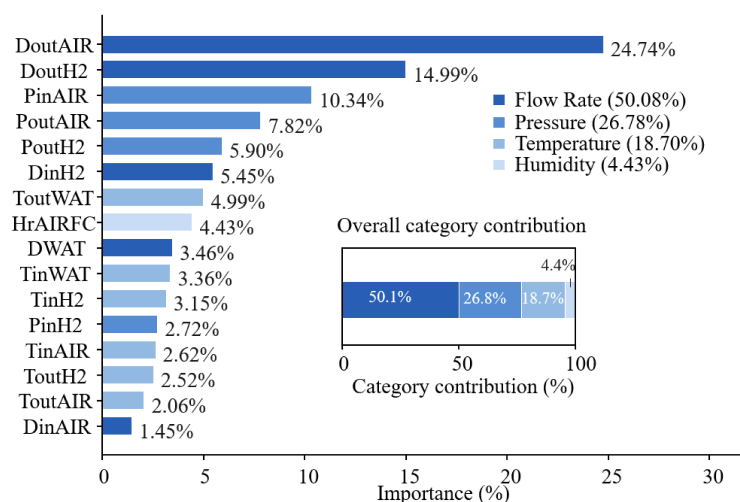


Figure 3. Importance ranking of operating parameters obtained by feature ablation analysis.

4.3. Feature Effect Direction and Sensitivity Analysis

Figure 4 illustrates the feature effect direction obtained by feature ablation analysis. Positive values indicate that replacing the corresponding feature with its reference value increases the predicted voltage, while negative values indicate a decrease in the predicted voltage. The outlet flow variables, especially $D_{out,AIR}$ and $D_{out,H2}$, exhibit the strongest negative effects, suggesting that their original values provide important degradation-related information for maintaining accurate voltage prediction.

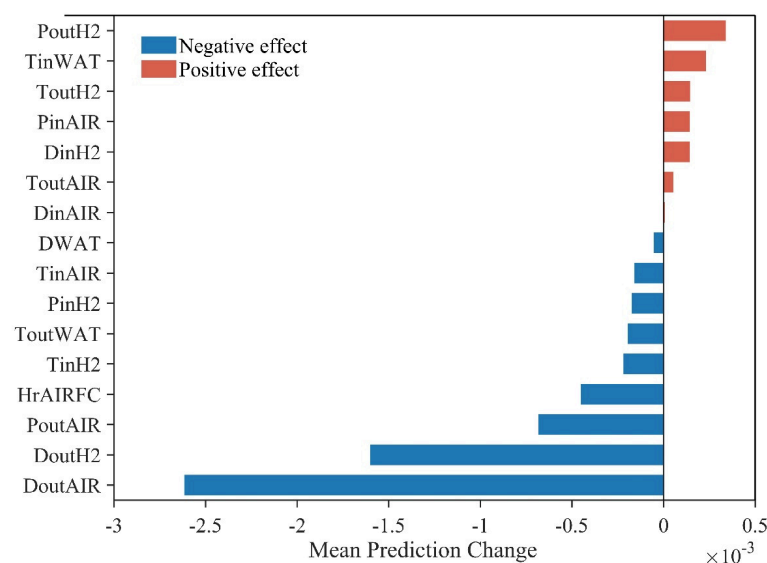


Figure 4. Characteristic effect direction.

To further examine the response of the Transformer model to different feature value levels, each operating variable was replaced by its 5th, 25th, 50th, 75th, and 95th percentile values from the training distribution. The resulting mean prediction changes are shown in Figure 5; positive values represent an upward shift in the predicted voltage, whereas negative values represent a downward shift.

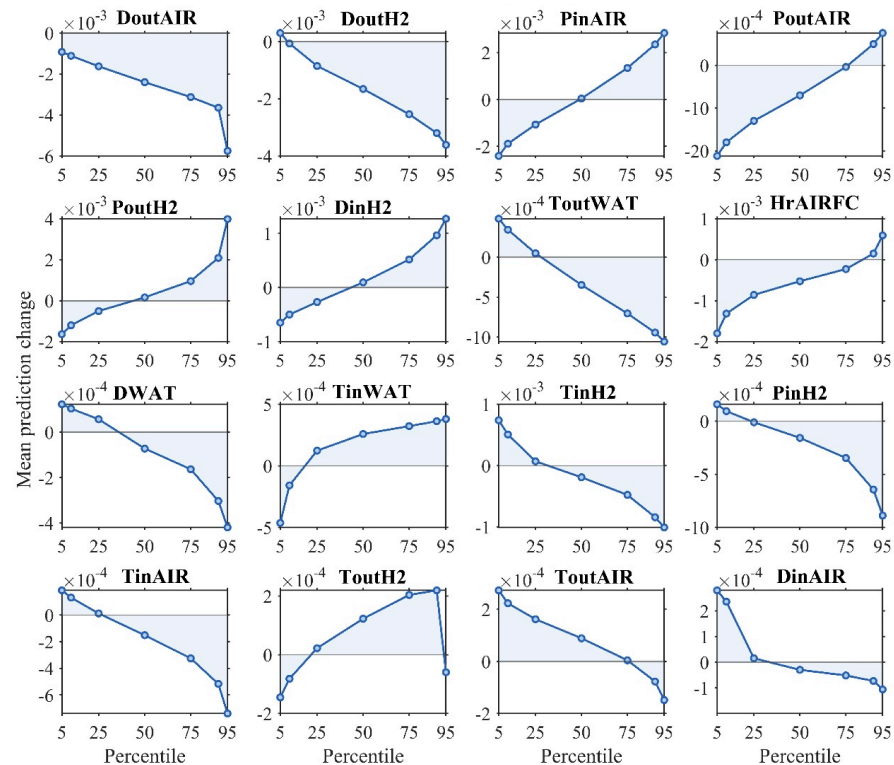


Figure 5. Feature sensitivity curves obtained by replacing each operating variable with different percentile values from the training distribution.

The sensitivity curves show that different variables have distinct response patterns. The outlet flow variables $D_{out,AIR}$ and $D_{out,H2}$ exhibit nearly monotonic decreasing trends as the replacement percentile increases, indicating that the model is highly sensitive to outlet gas flow information. This result further confirms the dominant role of mass transport-related variables in voltage degradation prediction under constant-current operation.

Pressure-related variables show more nonlinear responses. For example, $P_{in,AIR}$ and $P_{out,H2}$ generally increase with the percentile level, whereas $P_{out,AIR}$ and $P_{in,H2}$ tend to produce stronger negative effects at higher percentiles. This suggests that pressure variables influence the prediction in a range-dependent manner rather than through a simple linear relationship.

Temperature and humidity variables show relatively weaker responses. Although some temperature variables exhibit clear increasing or decreasing trends, their overall magnitudes are smaller than those of the main flow and pressure variables. The humidity variable $H_{r,AIRFC}$ also shows a mild effect, indicating that water management information contributes to the prediction but is not the dominant factor under the FC1 constant-current condition.

Overall, the sensitivity analysis supports the global attribution results. The Transformer model is most sensitive to outlet flow and pressure-related variables, while temperature and humidity variables mainly provide auxiliary correction information. This indicates that the model has learned degradation-related patterns closely associated with PEMFC mass transport behavior.

4.4. Relative Temporal-Position Sensitivity Analysis

To determine which historical positions within the input window influence the Transformer prediction, relative temporal-position sensitivity was evaluated through feature ablation. Figure 6 shows the normalized importance of the 50 relative temporal positions from $t - 49$ to t within the input window. These positions describe where an observation occurs in the historical input sequence and do not represent single-step or multi-step forecasting horizons. Because contributions at earlier historical positions are close to zero, the recent portion is emphasized for readability.

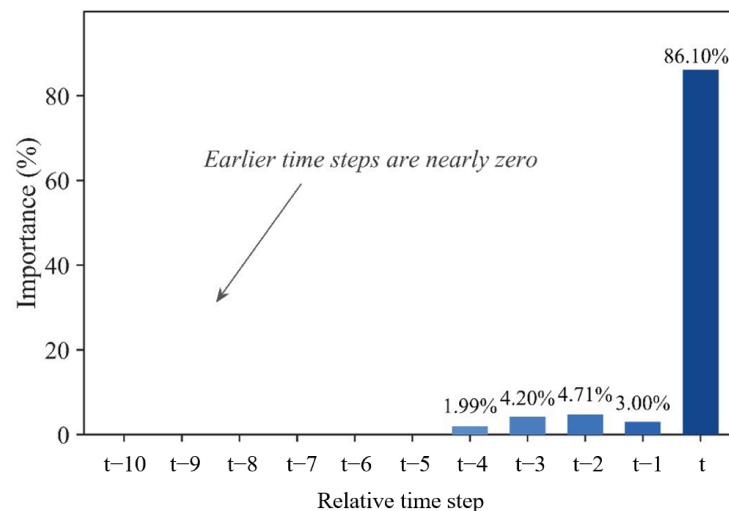


Figure 6. Importance distribution across relative temporal positions within the historical input window, obtained by feature ablation analysis.

As shown in Figure 6, temporal importance is strongly concentrated at the most recent relative positions within the input window. Position t contributes 86.10% of the total temporal importance. Positions $t - 1$, $t - 2$, $t - 3$, and $t - 4$ contribute 3.00%, 4.71%, 4.20%, and 1.99%, respectively, while positions farther back in the historical window provide negligible contributions.

This result indicates that distant historical observations provide limited marginal contributions under the present FC1 constant-current condition. Under slowly varying constant-current operation, the latest voltage and sensor measurements already provide a compact representation of the current degradation state and therefore dominate the corresponding voltage estimate. From a modeling perspective, this finding illustrates adaptive temporal information selection: although the complete 50-step window is available, the Transformer assigns relevance primarily to the observations that are most informative for this operating condition. The temporal distribution may differ under more dynamic operation and should be evaluated separately.

5. Conclusions

This study developed an interpretable Transformer-based framework for PEMFC voltage degradation prediction under constant-current operation. The FC1 sequence was processed using outlier correction, Savitzky–Golay filtering, normalization, and sliding-window reconstruction. The Transformer, RNN, LSTM, and GRU models were then evaluated using the same chronological sliding-window protocol over the held-out late-life interval. The Transformer achieved the best performance, with MAE of 8.2×10^{-4} , RMSE of 1.18×10^{-3} , MAPE of 0.0256%, and R^2 of 0.9961.

Feature ablation analysis was further conducted to interpret the prediction mechanism. The results indicate that flow-related variables dominate the model decision process, contributing 50.08% of the total importance, followed by pressure-related variables with 26.78%. Among all operating parameters, the air outlet flow rate $D_{\text{out,AIR}}$ and hydrogen outlet flow rate $D_{\text{out,H2}}$ showed the highest contributions, suggesting that outlet gas transport information is highly relevant to steady-state voltage degradation. Relative temporal-position sensitivity analysis showed that position t within the historical input window contributed 86.10% of the total temporal importance. This result indicates that, under the slowly varying FC1 constant-current condition, the model relies primarily on the most recent operating states within the input sequence.

Overall, the proposed framework provides both accurate prediction and physically meaningful interpretation for steady-state PEMFC degradation. The attribution results are consistent with the mass transport characteristics of PEMFC aging, demonstrating the potential of combining Transformer-based modeling with feature ablation analysis for interpretable fuel cell health prognostics.

Author Contributions: Conceptualization, F.Y., X.S., H.G. and J.Z. (Jiaming Zhou); methodology, F.Y. and X.S.; software, F.Y.; validation, F.Y., X.S., J.Z. (Jinming Zhang). and Z.H.; formal analysis, F.Y. and X.S.; investigation, F.Y., J.Z. (Jiaming Zhou) and J.Z. (Jinming Zhang); resources, H.G., S.W., X.L. and J.Z. (Junling Zhang); data curation, F.Y. and X.S.; writing—original draft preparation, F.Y.; writing—review and editing, X.S., H.G., S.W., X.L. and J.Z. (Junling Zhang); visualization, F.Y. and Z.H.; supervision, J.Z. (Jiaming Zhou); project administration, H.G. and J.Z. (Jiaming Zhou); funding acquisition, J.Z. (Jinming Zhang). All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by Weifang City Science and Technology Development Plan Projects (College and University Section) (2025GX037).

Data Availability Statement: The FC1 fuel-cell ageing data analyzed in this study are publicly available as part of the IEEE PHM 2014 Data Challenge dataset, provided by FCLAB and archived in the dataUBFC repository, at <https://doi.org/10.25666/DATAUBFC-2021-07-19>. No new experimental data were generated in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jia, C.; Liu, W.; Chau, K.; He, H.; Zhou, J.; Niu, S. Passenger-aware reinforcement learning for efficient and robust energy management of fuel cell buses. *eTransportation* **2026**, *27*, 100537. [\[CrossRef\]](#)
2. Lu, D.; Yi, F.; Hu, D.; Li, J.; Yang, Q.; Wang, J. Online optimization of energy management strategy for FCV control parameters considering dual power source lifespan decay synergy. *Appl. Energy* **2023**, *348*, 121516. [\[CrossRef\]](#)
3. 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*, 18534–18545. [\[CrossRef\]](#)
4. Jia, C.; He, H.; Zhou, J.; Li, J.; Wei, Z.; Li, K.; Li, M. A novel deep reinforcement learning-based predictive energy management for fuel cell buses integrating speed and passenger prediction. *Int. J. Hydrogen Energy* **2025**, *100*, 456–465. [\[CrossRef\]](#)
5. Ma, J.; Zhou, Y.; Lian, J. Charging the battery via an electric machine fed by dual power source in fuel cell vehicles. *J. Energy Storage* **2024**, *75*, 109700. [\[CrossRef\]](#)
6. Tao, J.; Zhang, Y.; Wei, X.; Jiang, S.; Dai, H. Optimization of fast cold start strategy for PEM fuel cell stack. *Appl. Energy* **2024**, *362*, 123018. [\[CrossRef\]](#)
7. Lu, D.; Hu, D.; Wang, J.; Wei, W.; Zhang, X. A Data-Driven Vehicle Speed Prediction Transfer Learning Method with Improved Adaptability Across Working Conditions for Intelligent Fuel Cell Vehicle. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 10881–10891. [\[CrossRef\]](#)
8. Nguyen, H.-L.; Lee, S.-M.; Yu, S. A comprehensive review of degradation prediction methods for an automotive proton exchange membrane fuel cell. *Energies* **2023**, *16*, 4772. [\[CrossRef\]](#)

9. Wang, Y.; Wu, K.; Zhao, H.; Li, J.; Sheng, X.; Yin, Y.; Du, Q.; Zu, B.; Han, L.; Jiao, K. Degradation prediction of proton exchange membrane fuel cell stack using semi-empirical and data-driven methods. *Energy AI* **2023**, *11*, 100205. [\[CrossRef\]](#)
10. Tang, X.; Qin, X.; Wei, K.; Xu, S. A novel online degradation model for proton exchange membrane fuel cell based on online transfer learning. *Int. J. Hydrogen Energy* **2023**, *48*, 13617–13632. [\[CrossRef\]](#)
11. Sun, B.; Liu, X.; Wang, J.; Wei, X.; Yuan, H.; Dai, H. Short-term performance degradation prediction of a commercial vehicle fuel cell system based on CNN and LSTM hybrid neural network. *Int. J. Hydrogen Energy* **2023**, *48*, 8613–8628. [\[CrossRef\]](#)
12. Zhang, F.; Ni, M.; Tai, S.; Zu, B.; Xi, F.; Shen, Y.; Wang, B.; Qin, Z.; Wang, R.; Guo, T.; et al. Machine learning assisted health status analysis and degradation prediction of aging proton exchange membrane fuel cells. *Appl. Energy* **2025**, *384*, 125483. [\[CrossRef\]](#)
13. Li, J. Research on PEMFC Water–Electricity–Magnetism Non-Uniform Distribution and Degradation Law. Master's Thesis, Southwest Jiaotong University, Chengdu, China, 2024.
14. Singh, R.; Sui, P.; Wong, K.; Kjeang, E.; Knights, S.; Djilali, N. Modeling the effect of chemical membrane degradation on PEMFC performance. *J. Electrochem. Soc.* **2018**, *165*, F3328–F3336. [\[CrossRef\]](#)
15. Jia, C.; He, H.; Zhou, J.; Li, K.; Li, J.; Wei, Z. A performance degradation prediction model for PEMFC based on bi-directional long short-term memory and multi-head self-attention mechanism. *Int. J. Hydrogen Energy* **2024**, *60*, 133–146. [\[CrossRef\]](#)
16. Ao, Y.; Laghrouche, S.; Depernet, D.; Chen, K. Proton exchange membrane fuel cell prognosis based on frequency-domain Kalman filter. *IEEE Trans. Transp. Electr.* **2021**, *7*, 2332–2343. [\[CrossRef\]](#)
17. Pei, P.; Chang, Q.; Tang, T. A quick evaluating method for automotive fuel cell lifetime. *Int. J. Hydrogen Energy* **2008**, *33*, 3829–3836. [\[CrossRef\]](#)
18. Meng, X.; Liu, M.; Mei, J.; Li, X.; Grigoriev, S.; Hasanien, H.M.; Tang, X.; Li, R.; Sun, C. Polarization loss decomposition-based online health state estimation for proton exchange membrane fuel cells. *Int. J. Hydrogen Energy* **2025**, *157*, 150162. [\[CrossRef\]](#)
19. Ma, R.; Yang, F.; Huo, Z.; Xie, R.; Zhao, D.; Gao, F. Multi-physical field modeling and model parameter sensitivity analysis of proton exchange membrane fuel cells. *Proc. CSEE* **2024**, *44*, 2699–2710.
20. Liu, M.; Meng, X.; He, T.; He, P.; Xu, K.; Yang, B. Intelligent parameter identification of proton exchange membrane fuel cells based on radial basis function neural network. *J. Kunming Univ. Sci. Technol. (Nat. Sci.)* **2024**, *49*, 81–90.
21. He, Y.; Qiao, L.; Zio, E.; Su, H.; Zhang, L.; Yang, Z.; Peng, S.; Zhang, J. A framework based on temporal causal inference graph neural networks for the probabilistic estimation of the remaining useful life of proton exchange membrane fuel cells. *Reliab. Eng. Syst. Saf.* **2025**, *265*, 111584. [\[CrossRef\]](#)
22. Jin, J.; Chen, Y.; Xie, C.; Wu, F. Degradation prediction of PEMFC based on data-driven method with adaptive fuzzy sampling. *IEEE Trans. Transp. Electr.* **2023**, *10*, 3363–3372. [\[CrossRef\]](#)
23. Xu, B.; Zhang, M.; Yang, W.; Zhang, L. Hybrid remaining useful life prediction method for proton exchange membrane fuel cells based on aging features. *Trans. China Electrotech. Soc.* **2025**, *40*, 5703–5714.
24. Zhao, B.; Zhang, L.; Zhang, L.; Chen, Z.; Liu, X.; Xie, C. Fusion prediction method for PEMFC lifetime based on joint iteration of TCN and AUKF. *Proc. CSEE* **2025**, *45*, 3609–3624.
25. Cheng, Y.; Zerhouni, N.; Lu, C. A hybrid remaining useful life prognostic method for proton exchange membrane fuel cell. *Int. J. Hydrogen Energy* **2018**, *43*, 12314–12327. [\[CrossRef\]](#)
26. Liu, H.; Chen, J.; Hissel, D.; Su, H. Remaining useful life estimation for proton exchange membrane fuel cells using a hybrid method. *Appl. Energy* **2019**, *237*, 910–919. [\[CrossRef\]](#)
27. Ke, C.; Han, K.; Wang, Y. A hybrid method combining degradation mechanisms and deep learning for lifetime prediction of proton exchange membrane fuel cells under dynamic load cycle conditions. *J. Power Sources* **2025**, *633*, 236464. [\[CrossRef\]](#)
28. Du, Q.; Zhan, Z.; Wen, X.; Zhang, H.; Tan, Y.; Li, S.; Pan, M. A hybrid model to assess the remaining useful life of proton exchange membrane fuel cells. *Processes* **2023**, *11*, 1583. [\[CrossRef\]](#)
29. Meng, X.; Sun, C.; Mei, J.; Tang, X.; Hasanien, H.M.; Jiang, J.; Fan, F.; Song, K. Fuel cell life prediction considering the recovery phenomenon of reversible voltage loss. *J. Power Sources* **2025**, *625*, 235634. [\[CrossRef\]](#)
30. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
31. Li, H.; Qiao, B.; Hou, Z.; Liu, J.; Yang, Y.; Lu, G. An interpretable data-driven method for degradation prediction of proton exchange membrane fuel cells based on temporal fusion transformer and covariates. *Int. J. Hydrogen Energy* **2023**, *48*, 25958–25971. [\[CrossRef\]](#)
32. Meng, X.; Mei, J.; Tang, X.; Jiang, J.; Sun, C.; Song, K. The degradation prediction of proton exchange membrane fuel cell performance based on a Transformer model. *Energies* **2024**, *17*, 3050. [\[CrossRef\]](#)
33. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Covert, I.; Lundberg, S.; Lee, S.I. Explaining by removing: A unified framework for model explanation. *J. Mach. Learn. Res.* **2021**, *22*, 209.

35. Gouriveau, R.; Hilaiet, M.; Hissel, D. IEEE PHM 2014 data challenge: Outline, experiments, scoring of results, winners. In Proceedings of the 2014 IEEE Conference on Prognostics and Health Management (PHM), Cheney, WA, USA, 22–25 June 2014.
36. Zhou, J.; Shu, X.; Zhang, J.; Yi, F.; Jia, C.; Zhang, C. A deep learning method based on CNN-BiGRU and attention mechanism for proton exchange membrane fuel cell performance degradation prediction. *Int. J. Hydrogen Energy* **2024**, *94*, 394–405. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.