Journal of Zhejiang University-SCIENCE A 2026 27(5):506-517
www.jzus.zju.edu.cn; www.springer.com/journal/11582
E-mail: jzus_a@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.A2500337

Check for updates

Real-time degradation modeling for automotive PEMFC stacks: a multi-scale fusion network validated on an industrial 215-channel system

Zifei WANG[1,2], Xiangxian ZHU[3], Congxin LI[4], Daidai CHEN[3], Zhitao LIU[2], Longhua MA[1], Jili TAO[1✉], Hongye SU[2]

[1]*School of Information Science and Engineering, NingboTech University, Ningbo 315100, China*
[2]*State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China*
[3]*Ningbo Joyson Electronic Corp., Ningbo 315040, China*
[4]*Ningbo Green Power Hydrogen Technology Research Institute Co., Ltd., Ningbo 315033, China*

Abstract: Accurately predicting long-term degradation patterns in proton exchange membrane fuel cell (PEMFC) stacks under automotive operating conditions remains challenging. Prediction methods are largely constrained by laboratory-scale experiments and limited stack sizes, resulting in insufficient accuracy and generalization capability. To address these limitations, in this paper we propose a multi-scale bidirectional fusion network (MBFNet) tailored for an industrial 215-channel PEMFC stack, enabling accurate degradation prediction under accelerated real-world dynamic conditions using gas-heat-electricity (GHE) co-simulation data. A channel-joint adaptive noise correlation threshold (NCT) algorithm is introduced to account for variable correlations across sensors and operating conditions without relying on prior physical modeling. A multi-scale decomposition module captures degradation dynamics at different temporal scales, while a bidirectional fusion module integrates global trends and local details into the final prediction. Experimental results show that MBFNet achieves 18.6% lower prediction error and 36.8% fewer parameters than the long short-term memory (LSTM)-attention benchmark under real operating scenarios. In multi-step prediction tasks, MBFNet reduces root mean square error by an average of 24.5% relative to LSTM-attention and 55.2% relative to a one-dimensional convolutional neural network (1D-CNN) across four prediction horizons, better satisfying automotive application requirements. Moreover, MBFNet exhibits strong physical interpretability, making it efficient to implement and promising for practical deployment.

Key words: Large-scale proton exchange membrane fuel cell (PEMFC) stacks; Multi-scale bidirectional fusion network (MBFNet); Degradation prediction; Automotive scenarios

1 Introduction

Proton exchange membrane fuel cells (PEMFCs) represent a pivotal technology for directly converting hydrogen's chemical energy into electricity, with strong potential for decarbonizing the transportation sector (Chen et al., 2022; Shin and Yoo, 2023). Compared with lithium batteries, PEMFCs offer zero emissions, rapid refueling, and excellent cold-start capability, making them attractive for range-demanding applications such as heavy-duty trucks and cold-chain logistics (He et al., 2024). However, commercialization remains constrained by durability issues. Under real automotive conditions, degradation is often 3–5 times faster than in controlled laboratory tests, substantially increasing lifecycle costs (Chen et al., 2019). This acceleration is driven by rapidly varying operating profiles—start/stop cycles, dynamic load fluctuations, and environmental changes—which induce complex failure mechanisms including catalyst layer detachment, membrane degradation, and bipolar plate corrosion (Ma et al., 2023; Tian et al., 2023). Therefore, accurate lifetime prediction is essential for effective prognostics and health management (PHM) and for

✉ Jili TAO, taojili@nbt.edu.cn
Jili TAO, https://orcid.org/0000-0001-7095-8968

Received July 22, 2025; Revision accepted Dec. 16, 2025;
Crosschecked Apr. 22, 2026; Online first May 18, 2026

improving durability while reducing costs in automotive applications (Liu et al., 2019).

Despite substantial progress, a notable gap persists between laboratory studies and engineering requirements, mainly due to three limitations. (1) Stack size: most studies focus on stacks with fewer than 50 cells, overlooking large-stack effects such as bipolar plate contact pressure heterogeneity and flow distribution fluctuations (Tian et al., 2025). This restricts understanding of multi-physics coupling in industrial-scale stacks (>200 cells), where degradation rates can vary by up to 30% (Benaggoune et al., 2022). (2) Operating condition bias: simplified laboratory protocols (e.g., constant/cyclic loads) fail to reflect environmental variability and dynamic-load-induced phenomena such as microchannel deformation during rapid current transients (Zuo et al., 2021). (3) Noise coupling: turbulent pulsation increases inlet-side sensor noise by from 20% to 40% relative to the outlet, and conventional denoising may reduce the effective signal mutual information by up to 15% (Benaggoune et al., 2022; Wang et al., 2023). Together with an overreliance on voltage/current monitoring that neglects coupled variables (e.g., pressure and temperature), these issues hinder multi-physics degradation modeling. Consequently, prediction errors in real scenarios can be 3–5 times higher than those of laboratory benchmarks (Zuo et al., 2021), highlighting the need for models that exploit high-fidelity and multi-variable data.

Efforts to predict PEMFC lifetime can be broadly categorized into model-driven and data-driven approaches, each with distinct principles and limitations. Model-driven methods, including physical-mechanistic, empirical, and semi-empirical models (Chen et al., 2022), provide strong interpretability by explicitly linking degradation behavior to physical or empirical mechanisms. Representative studies have identified key failure pathways under automotive conditions, including membrane hydrogen crossover, catalyst layer structural degradation, membrane radical attack, Pt catalyst degradation, and carbon corrosion (Liu et al., 2014; Ren et al., 2020). To improve prediction accuracy under uncertainty, filtering-based techniques have been widely adopted. Wang et al. (2021) combined polarization resistance modeling with particle filtering to enhance generality and accuracy, while Jouin et al. (2014) and Bressel et al. (2016) used particle filtering and extended Kalman filtering, respectively, to dynamically estimate degradation trends and time-varying parameters. Although these approaches have been validated through long-term durability tests, model-driven methods still face challenges in parameter identification and computational scalability when addressing strong multi-physics coupling and material degradation in automotive environments.

In contrast, data-driven approaches bypass explicit mechanistic modeling and directly learn degradation patterns from operational data, spanning probabilistic frameworks and machine learning algorithms. Traditional machine learning models have been widely applied, including artificial neural networks (ANNs) for small-scale PEMFCs (Wilberforce and Olabi, 2021) and relevance vector machine (RVM) -based degradation prediction for limited training datasets (Wu et al., 2016). To improve computational efficiency, Zhou et al. (2022) showed that extreme learning machines (ELMs) can accelerate training while detecting degradation from basic sensor signals. Despite these advantages, data-driven methods generally require large, high-quality datasets and remain sensitive to noise and sensor drift, while offering limited physical interpretability. To mitigate these limitations, hybrid-driven approaches have emerged. Zhou et al. (2017) integrated multi-physics aging models with particle filtering to incorporate mass transfer and ohmic loss mechanisms. Cheng et al. (2018) combined operational data with empirical models via unscented Kalman filtering to enhance robustness under automotive conditions.

Deep learning methods for PEMFC lifetime prediction have increasingly focused on recurrent neural networks (RNNs) and their variants due to their suitability for time-series modeling. Early advances used enhanced recurrent architectures to improve temporal dependency learning. Ma et al. (2018) introduced grid long short-term memory (G-LSTM) networks for PEMFC degradation prediction, achieving less than 5% error across 1.2–25.0 kW stacks under multiple operating conditions while enabling online deployment. Yi et al. (2025) further proposed matrix-LSTM, which leverages exponential gating and matrix memory to address long-range dependence in high-dimensional degradation data while supporting parallel computation. Nevertheless, standard RNN-based models suffer from three inherent limitations: increasing computational burden and error accumulation with longer sequences, insufficient separation of recoverable dynamics from

irreversible degradation, and limited capability to capture multivariate feature correlations.

To overcome these challenges, attention mechanisms and hybrid deep learning architectures have been introduced. He et al. (2022) incorporated explicit feature engineering by extracting degradation-sensitive health indicators using auto-encoders coupled with LSTM to improve prediction accuracy and enable degradation mechanism analysis. Li et al. (2023) adopted a decomposition-based strategy, separating voltage degradation into multi-scale components and modeling linear trends and nonlinear recovery dynamics using an autoregressive integrated moving average (ARIMA) and an attention-enhanced gated recurrent unit (GRU), respectively. Beyond RNN-based approaches, computer vision-inspired architectures have also been explored. Liu et al. (2022) proposed Inc-DenseNet, which converts voltage signals into two-dimensional representations to distinguish flooding and air starvation faults with high diagnostic accuracy. Despite these advances, deep learning models still face persistent bottlenecks, including limited generalization under highly dynamic automotive conditions, reliance on manual hyperparameter tuning (Sahajpal et al., 2023), and excessive computational complexity. Attention mechanisms can cause exponential parameter growth, leading to memory overload and inference latencies exceeding 500 ms on automotive electronic control units (ECUs) (Wilberforce et al., 2023; Sun et al., 2024). For large-scale PEMFC stacks with hundreds of parallel channels, such computational demands often exceed onboard resource constraints, driving the development of lightweight, edge-compatible architectures (Xia et al., 2022).

Overall, despite significant progress, deep learning models still face four intertwined bottlenecks that limit their practical deployment for industrial-scale PEMFC stacks: (1) ineffective fusion of cross-scale temporal features (e.g., transient recovery versus long-term aging), (2) a lack of adaptive optimization beyond manual hyper-parameter tuning, (3) insufficient embedding of physical mechanisms to enhance interpretability and generalization, and (4) prohibitive computational complexity that fails to meet onboard resource constraints. To directly address these challenges, the main contributions of this study are summarized as follows:

1. Establishment of an industrial-scale gas-heat-electricity (GHE) co-simulation platform. A unique degradation dataset is collected from a 215-channel PEMFC stack under accelerated east China climatic conditions. The platform enables synchronized and high-resolution monitoring of flow field pulsation and electrochemical parameters, effectively mitigating data biases caused by distorted laboratory conditions and insufficient stack scale.

2. Development of a channel-joint adaptive noise correlation threshold (NCT) algorithm. A dynamic denoising method is proposed to construct spatio-temporal noise thresholds through dual-channel coherence analysis without relying on prior physical modeling. The NCT algorithm explicitly addresses coupled noise interference across multi-physical sensor signals, improving data quality for subsequent modeling.

3. Proposal of the multi-scale bidirectional fusion network (MBFNet). A neural architecture is introduced to overcome the limitations of existing models. It consists of (1) a multi-scale decomposition module that disentangles sensor couplings to capture degradation dynamics at different timescales (e.g., short-term voltage recovery versus long-term aging) and (2) a bidirectional fusion module that uses parameter-efficient projections to reconcile stack-level degradation trends with cell-level anomalies, ensuring both accuracy and computational efficiency.

4. Design of a rigorous aging prediction validation framework. A comprehensive experimental protocol is implemented to demonstrate significantly improved prediction accuracy under dynamic operating conditions, with particular emphasis on the critical degradation phase. This framework provides a validated technical paradigm for predictive health management in fuel cell vehicles, bridging the gap between algorithm innovation and automotive applications.

2 Dataset

2.1 Experimental system

The dataset used in this study was obtained from accelerated aging tests conducted on an industrial-grade GHE co-simulation platform based on a 215-channel PEMFC stack with a rated power of 95 kW. The platform was designed to emulate realistic automotive operating conditions, including dynamic load cycling, frequent start–stop events, and coupled electro-thermal-fluidic stresses, which are known to significantly

accelerate degradation in large-scale stacks. The PEMFC stack consists of 430 individual fuel cells electrically configured into 215 series-connected cell pairs for voltage monitoring. Throughout this paper, "cell voltage" refers to the voltage measured across each paired channel unit. The system provides synchronized and high-resolution measurements of voltage, current, pressure, temperature, humidity, and flow rate, with spatio-temporal alignment errors below 5 ms, enabling reliable characterization of degradation behavior under complex operating conditions. A total of 327 h of operational data were collected under harsh automotive scenarios representative of east China. The end-of-life (EOL) criterion is defined as a 10.5% reduction in stack output voltage relative to the initial peak operating voltage (from 285.0 to 255.0 V). Linear regression analysis over the full aging period reveals a statistically significant degradation trend (R^2= 0.871), with the voltage decreasing from 280.6 to 266.3 V, corresponding to a total decay of 14.3 V and an average degradation rate of 203.42 μV/(cell·h). This rate is about 2.91 times higher than that reported under standardized drive cycles (Liu et al., 2014), reflecting the intentionally aggressive testing protocol adopted to accelerate aging and evaluate model robustness under extreme conditions. Actual vehicle durability under normal operating conditions is therefore expected to be higher.

As a result, the dataset exhibits complex degradation trajectories characterized by pseudo-cyclic fluctuations and abrupt disturbances induced by environmental interactions, encompassing both short-term transient responses and long-term irreversible aging behavior. Detailed descriptions of the experimental platform, onsite testing configuration, and degradation trend visualization are provided in Section S1 of the electronic supplementary materials (ESM).

2.2 Data preprocessing and validation

Considering the strong non-stationarity and multivariate coupling inherent in automotive PEMFC operational data, a three-step preprocessing framework was designed to extract degradation-relevant features.

2.2.1 Condition focusing and downsampling

The analysis focuses on data collected under the nominal operating current of (330±5) A, which accounts for about 72% of the total operational duration. Data segments associated with atypical operating events, such as system start–stop transients and fault conditions, are excluded. The original 5-Hz signals are downsampled to 0.2 Hz using moving-average aggregation over 5-s windows, achieving a balance between computational efficiency and dynamic feature preservation.

2.2.2 Noise suppression and feature selection

To mitigate signal distortion and enhance degradation feature extraction, a systematic preprocessing pipeline was implemented, comprising four complementary steps.

1. Environmental noise suppression: sliding-window normalization with a 120-s window was applied to temperature and humidity signals to suppress low-frequency environmental disturbances.

2. Electrical stack noise processing: a channel-joint adaptive NCT algorithm was used to dynamically suppress coupled sensor noise based on dual-channel coherence analysis (Section 3).

3. Feature selection: seven variables exhibiting strong nonlinear correlation with voltage degradation (mutual information>0.15) were selected, including hydrogen inlet temperature ($TinH_2$), air activation temperature (TactAIR), air outlet temperature (ToutAIR), relative humidity (RH), air inlet pressure (PinAIR), air outlet pressure (PoutAIR), and air flow rate (QAIR) (Fig. 1). This reduces the input dimensionality by 42% while preserving physically meaningful degradation-related features. The detailed formulation of mutual information is provided in Section S3 of the ESM.

4. Normalization: all variables were scaled to the [0, 1] interval using min–max normalization to account for differences in physical units and magnitudes.

2.2.3 Dataset partitioning and validation

The first 229-h data were used for model training, covering normal operation and early-stage degradation, while the remaining 98-h data were reserved for testing to evaluate predictive performance under accelerated aging conditions. To emulate real-time deployment, a sliding-window strategy was adopted with a 10-min input sequence and a 1-min prediction horizon. The overall voltage evolution, dataset partitioning, and statistical distribution characteristics are provided in Section S1 of the ESM.

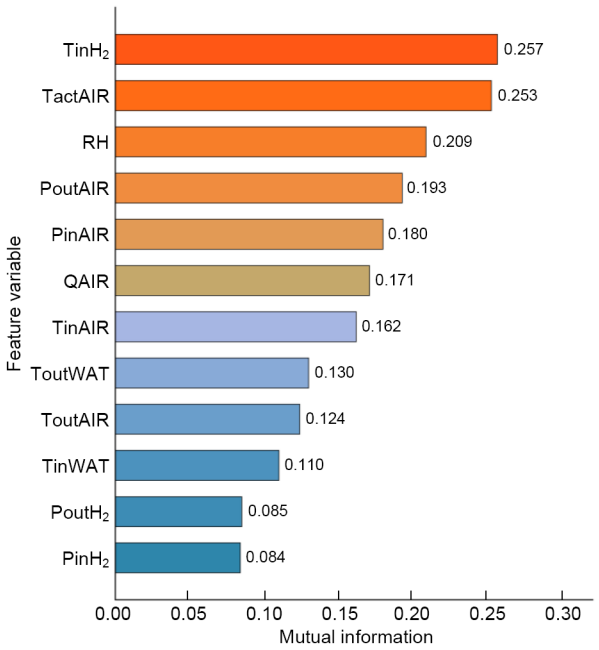

Fig. 1 Mutual information scores for variables (ToutWAT: temperature out of water; TinWAT: temperature of inlet water)

3 Multi-scale bidirectional fusion network

The degradation dynamics of large-scale PEMFC systems under complex operating conditions (e.g., dynamic loading and start–stop cycles) exhibit intricate electro-thermal-fluidic coupling, multi-time scale interactions, and multi-source sensor noise interference. Conventional methods struggle to simultaneously model these coupled phenomena, often exhibiting significant prediction bias. To address these complexities, we propose the MBFNet, comprising: (1) a channel-joint adaptive NCT algorithm for coupled noise suppression, (2) multi-scale decomposition for disentangling transient/long-term features, and (3) parameter-efficient bidirectional fusion reconciling stack-level trends with cell-level degradation.

3.1 Adaptive noise correlation threshold

Sensor signals in automotive PEMFC systems are strongly affected by turbulent pulsation and dynamic operating conditions. Inlet-side measurements typically exhibit noise energy that is from 20% to 40% higher than that at the outlet side. Under rapid load transients, the noise power spectral density (PSD) of different sensor channels becomes time-varying and mutually correlated, which can induce oscillatory distortion in conventional denoising schemes. Moreover, noise components in physically coupled variables—such as temperature and pressure—often exhibit coherence, causing traditional frequency-domain filtering to suppress useful information and reduce mutual information in the effective signal. To address these challenges, a channel-joint adaptive NCT algorithm is proposed to explicitly exploit the coherence between physically correlated sensor channels (e.g., inlet and outlet pressures and co-located temperature sensors). Instead of treating each signal independently, the NCT algorithm dynamically estimates a noise threshold by modeling the coherence behavior between paired channels, enabling adaptive discrimination between correlated signal components and uncorrelated noise. This design is particularly effective for PEMFC systems operating under start–stop cycles and dynamic load variations, where multi-physical noise coupling is prominent.

The NCT algorithm operates on framed dual-channel signals and combines spectral coherence estimation with adaptive thresholding to track transient noise characteristics. By incorporating smoothed auto- and cross-power spectral densities, the algorithm captures time-varying coherence propagation during operational transients. A tentative noise correlation threshold is then derived from the discrepancy between modeled coherence under estimated noise conditions and the observed coherence, allowing the denoising strength to adapt automatically to load-induced noise amplification. To ensure numerical stability and prevent excessive suppression of degradation-related features, the threshold is further bounded and temporally smoothed across operating cycles. The resulting frame-level NCT is applied as a time-varying threshold in the wavelet domain, enabling adaptive denoising with minimal phase delay while preserving critical degradation information. Compared with fixed-threshold or single-channel filtering methods, the proposed NCT algorithm is fully data-driven and requires no prior physical noise modeling, making it robust to the complex and non-stationary noise characteristics encountered in large-scale automotive PEMFC systems. The detailed mathematical formulation of the NCT algorithm, including spectral estimation, coherence modeling, threshold construction, and wavelet-domain implementation, is provided in Section S2 of the ESM.

3.2 Multi-scale decomposition

When a large PEMFC stack is suddenly shut down after high load operation, processes such as redistribution of water molecules and regeneration of active sites on the catalyst surface will trigger the phenomenon of non-steady state voltage recovery. Degradation is nonlinear and can be related on different time scales, and traditional methods cannot effectively handle time lags and multi-scale features. The decomposition module specifically targets voltage recovery transients and aging drifts in large stacks by segmenting sequences into various scales: high-resolution series preserve rapid fluctuations (e.g., start–stop anomalies), while low-resolution series encapsulate slow degradation trends (e.g., catalyst decay).

Given input data $\boldsymbol{X}\in\mathbb{R}^{I\times F}$, decompose $\boldsymbol{X}$ with the average pooling as $\boldsymbol{X}_{\mathrm{E}}=\mathrm{Decomp}(\boldsymbol{X})$ to get M subsequences, where $\boldsymbol{X}_{\mathrm{E}}=\{\boldsymbol{X}_0, \boldsymbol{X}_1, \cdots, \boldsymbol{X}_m\}$, $m\in\{0, 1, \cdots, M\}$ and $\boldsymbol{X}_m\in\mathbb{R}^{\frac{I}{2^m}\times F}$ denotes the subsequence achieved by average pooling along the time axis. The decomposition module Decomp($\boldsymbol{X}$) uses average pooling with a kernel size of 2 and a stride of 2 along the temporal axis, progressively halving the sequence length. The number of decomposition scales M is set to 5, generating subsequences at temporal resolutions of [I, I/2, I/4, I/8, I/16]. This multi-scale structure is crucial for PEMFC data, as it explicitly separates short-term voltage recovery transients (captured in high-resolution scales) from long-term aging trends (captured in low-resolution scales). To fuse sensor data such as temperature, pressure, and flow, an encoder is designed to extend the dimensions of $\boldsymbol{X}_{\mathrm{E}}$ as $\boldsymbol{X}_{\mathrm{E}}^0=\mathrm{Encoding}(\boldsymbol{X}_{\mathrm{E}})$, which enhances the model's ability to characterize the complex processes inside power reactors. The size of each element in $\boldsymbol{X}_{\mathrm{E}}^0=\{\boldsymbol{X}_0^0, \boldsymbol{X}_1^0, \cdots, \boldsymbol{X}_m^0\}$ is $\mathbb{R}^{\frac{I}{2^m}\times F_{\mathrm{enc}}}$ to learn a nonlinear mapping, denoting the set of encoded sequences. The encoder Encoding ($\boldsymbol{X}_{\mathrm{E}}$) (is implemented as a two-layer fully connected network with rectified linear unit (ReLU) activation, which projects the feature dimension F to an encoded dimension F_{enc}=64, enhancing the model's capacity to represent complex intra-stack processes.

3.3 Bidirectional fusion

Large PEMFC stacks face three key challenges under dynamic operating conditions. (1) Limitations in temporal modeling: although recurrent models such as LSTM can retain temporal information through memory units, their sequential propagation causes early-state information to decay with increasing sequence length, limiting their ability to capture long-term degradation patterns. Attention mechanisms partially alleviate this issue by reweighting temporal features, but the resulting parameter growth often leads to overfitting and violates the inherent continuity of PEMFC aging. (2) Multi-scale degradation representation: frequent start-stop operations induce voltage recovery and irreversible degradation. Single-scale models face an inherent trade-off—models emphasizing local fluctuations are sensitive to operating disturbances, whereas those focusing on global trends fail to capture abrupt degradation events. (3) Computational burden: attention-enhanced recurrent models incur high latency due to large parameter counts and sequential iterations, making them unsuitable for real-time automotive deployment.

To address these challenges, the proposed fusion module decomposes spatio-temporal features across multiple scales and integrates them through an efficient bidirectional fusion mechanism, as illustrated below. Decompose the encoded sequence $\boldsymbol{X}_{\mathrm{E}}^0$ and get $\boldsymbol{X}_{\mathrm{ET}}^0, \boldsymbol{X}_{\mathrm{EL}}^0\in\mathbb{R}^{\frac{I}{2^m}\times F_{\mathrm{enc}}}$ as in Fig. 2:

$$\begin{aligned}\boldsymbol{X}_{\mathrm{ET}}^0&=\mathrm{mean}\left(\mathrm{AvgPool}(\mathrm{Padding}(\boldsymbol{X}_{\mathrm{E}}^0))_{k_n^{\mathrm{kernel}}}\right),\\ \boldsymbol{X}_{\mathrm{EL}}^0&=\boldsymbol{X}_{\mathrm{E}}^0-\boldsymbol{X}_{\mathrm{ET}}^0.\end{aligned}\tag{1}$$

Here, $\boldsymbol{X}_{\mathrm{ET}}^0$ characterizes overall stack degradation (e.g., bipolar plate corrosion and membrane electrode aging); $\boldsymbol{X}_{\mathrm{EL}}^0$ captures localized fluctuations at the cell level (e.g., diffusion layer flooding and sudden changes in contact resistance). AvgPool()$_{k_n^{\mathrm{kernel}}}$ represents the average pooling with multiple kernels k_n^{kernel}. Padding() means when samples have different lengths, zeros are used to pad them to the same length.

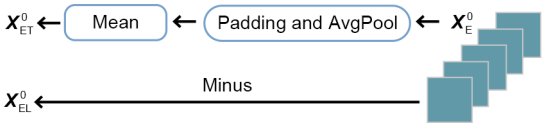

Fig. 2 Internal decomposition process

Since there are M groups of sequences in $\boldsymbol{X}_{\mathrm{ET}}^0=\{\boldsymbol{X}_{\mathrm{ET1}}^0, \boldsymbol{X}_{\mathrm{ET2}}^0, \cdots, \boldsymbol{X}_{\mathrm{ET}M}^0\}$, the information is extracted with macro-information from $m=M-1$ to $m=0$,

which means establishing a degradation baseline by first analyzing long-term trends at coarse resolutions (where system-level aging patterns dominate), and then incrementally refines predictions using finer temporal details.

$$\boldsymbol{X}^0_{\mathrm{ET}m} \leftarrow \boldsymbol{X}^0_{\mathrm{ET}m} + \mathcal{F}_{\mathrm{trend}}(\boldsymbol{X}^0_{\mathrm{ET}(m+1)}). \tag{2}$$

Similarly, there are M sets of sequences in $\boldsymbol{X}^0_{\mathrm{EL}} = \{\boldsymbol{X}^0_{\mathrm{EL}1}, \boldsymbol{X}^0_{\mathrm{EL}2}, \cdots, \boldsymbol{X}^0_{\mathrm{EL}M}\}$, and the micro-information from $m=1$ to M is used to detect localized anomalies by prioritizing high-resolution signals and then progressively incorporates broader contextual patterns to distinguish significant information from noise.

$$\boldsymbol{X}^0_{\mathrm{EL}m} \leftarrow \boldsymbol{X}^0_{\mathrm{EL}m} + \mathcal{F}_{\mathrm{local}}(\boldsymbol{X}^0_{\mathrm{EL}(m-1)}). \tag{3}$$

Here, $\mathcal{F}_{\mathrm{trend}}()$ and $\mathcal{F}_{\mathrm{local}}()$ are implemented as linear layers to simplify the model, with reverse dimension change between $\mathbb{R}^{\frac{I}{2^{m+1}}\times F_{\mathrm{enc}}}$ and $\mathbb{R}^{\frac{I}{2^m}\times F_{\mathrm{enc}}}$. Computing complexity grows only linearly with stack size and supports parallel processing of cells. The complete fusion is shown in Fig. 3.

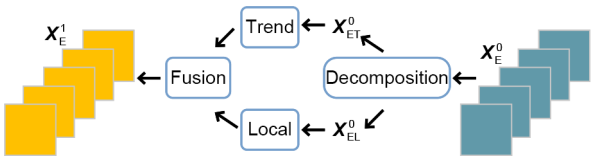

Fig. 3 Completed fusion process of $\boldsymbol{X}^0_{\mathrm{E}}$ to $\boldsymbol{X}^1_{\mathrm{E}}$

$$\boldsymbol{X}^1_{\mathrm{E}} = \boldsymbol{X}^0_{\mathrm{E}} + \boldsymbol{W} \otimes (\boldsymbol{X}^0_{\mathrm{ET}} \oplus \boldsymbol{X}^0_{\mathrm{EL}}), \tag{4}$$

where $\boldsymbol{W}$ is the learnable weight matrix, $\oplus$ denotes feature concatenation along the feature dimension, and $\otimes$ denotes the matrix multiplication operation (i.e., the linear transformation).

If multiple layers are required, $\boldsymbol{W}$ can be chosen as $\{\boldsymbol{W}_1, \boldsymbol{W}_2, \cdots \boldsymbol{W}_K\}$, and K can be set as 4. $\boldsymbol{X}^1_{\mathrm{E}}$ continues the cycle to get $\boldsymbol{X}^2_{\mathrm{E}}$.

Outputs $\boldsymbol{X}^1_{\mathrm{E}}, \boldsymbol{X}^2_{\mathrm{E}} \in \mathbb{R}^{\frac{I}{2^m}\times F_{\mathrm{enc}}}$ spanning scales $m \in \{0, 1, \cdots, M\}$ undergo multi-scale fusion through scale-specific linear projections $\mathcal{P}_m(\cdot)$ followed by residual overlays, formalized as Eq. (5). The voltage feature channel V_{index} then integrates fused representations across all scales via Eq. (6), breaking traditional single-scale fusion paradigms to achieve equilibrium of multi-scale degradation features within computational constraints by combining adaptive resolution alignment via $\mathcal{P}_m(\cdot)$ and feature preservation through residual connections $\oplus$. Here, $\boldsymbol{Y}_n$ is the nth group of output of MBFNet, and N is the number of $\boldsymbol{Y}_n$.

$$\boldsymbol{Y}_n = \mathcal{P}_m(\boldsymbol{X}^1_m \oplus \boldsymbol{X}^2_m),\ \forall m \in \{0,\ 1, \cdots,\ M\}, \tag{5}$$

$$\boldsymbol{Y}_{\mathrm{out}} = \sum_{n=0}^{N} \boldsymbol{Y}_n[:,\ V_{\mathrm{index}}]. \tag{6}$$

4 Case and experiments

4.1 Performance metrics

Prediction accuracy was quantified via root mean square error (E_{RMS}) and mean square error (E_{MS}), computed as the square root of the average squared deviations between predicted (p_n^{pre}) and actual (p_n^{real}) values. Similarly, mean absolute error (E_{MA}) measures the arithmetic mean of absolute deviations, and the R^2 coefficient describes the distribution of data and $\bar{p}_n$ denotes the mean values.

$$E_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(p_n^{\mathrm{pre}} - p_n^{\mathrm{real}})^2}, \tag{7}$$

$$E_{\mathrm{MA}} = \frac{1}{N}\sum_{n=1}^{N}|p_n^{\mathrm{pre}} - p_n^{\mathrm{real}}|, \tag{8}$$

$$E_{\mathrm{MS}} = \frac{1}{N}\sum_{n=1}^{N}(p_n^{\mathrm{pre}} - p_n^{\mathrm{real}})^2, \tag{9}$$

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(p_n^{\mathrm{pre}} - p_n^{\mathrm{real}})^2}{\sum_{n=1}^{N}(\bar{p}_n - p_n^{\mathrm{real}})^2}. \tag{10}$$

The selection principles of baseline models and the detailed architectural configuration of MBFNet, including layer-wise parameters and deployment constraints, are provided in Section S4 of the ESM.

4.2 Results and comparisons

Time-series features were extracted by using the rolling time windows, and all preprocessing operations were independently initialized using the training data. To verify the effectiveness of the MBFNet, in this experiment we compared and analyzed the performance metrics of six temporal prediction models under the same platform and dataset partitioning strategy, including the baseline models (LSTM, GRU, and

one-dimensional convolutional neural network (1D-CNN)) and the attention-enhanced variants (LSTM-attention and GRU-attention) (Table 1). All experiments were run on an NVIDIA RTX 4060 GPU, Intel® i7-13650HX 2.60 GHz CPU.

The single-step prediction performance of MBFNet and other models is shown in Fig. 4. Computationally, MBFNet shows a significant advantage in the balance of accuracy and efficiency. The prediction accuracy is optimized with an E_{RMS}=0.0180, which was 18.6% lower than that of LSTM-attention, and the mean absolute percentage error (E_{MAP}) index is 27.2% lower. The training time was 15.2 min, which was only 70.0% of that of LSTM-attention. The number of parameters in our model is 2.4×10^4, which is 36.8% less than that of LSTM-attention. The decomposition computation can be implemented in parallel for GPU acceleration, with a 32% speedup over serial implementation. In terms of the trade-off between accuracy and efficiency, although 1D-CNN had the shortest training time of 4.5 min, its E_{RMS} was 136% higher than that of MBFNet, suggesting that using convolution

Table 1 Single-step prediction performance

Model	R^2	E_{RMS}	E_{MA}	E_{MAP} (%)	Training time (min)	Size of parameters ($\times10^4$)
MBFNet (ours)	0.9654	0.0180	0.0123	0.3002	15.2	2.4
LSTM-attention	0.9482	0.0221	0.0157	0.4125	21.7	3.8
GRU-attention	0.9365	0.0253	0.0179	0.4983	18.9	3.6
LSTM	0.9124	0.0317	0.0214	0.6541	11.3	3.2
GRU	0.9038	0.0348	0.0236	0.7329	9.8	2.9
1D-CNN	0.8741	0.0425	0.0291	0.9257	4.5	1.7

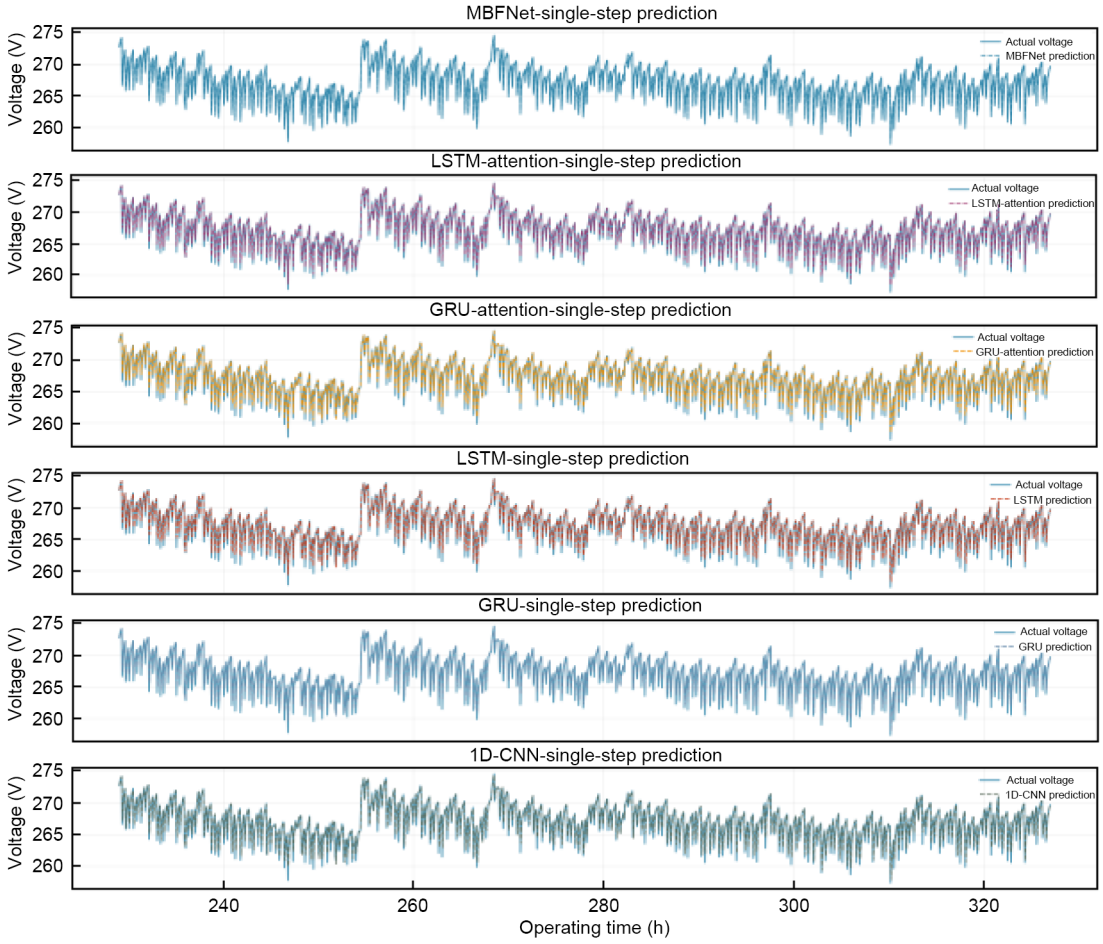

Fig. 4 Single-step prediction results of each model

only it is difficult to capture cross-scale correlations in the degradation process.

Table 2 shows the error accumulation under multi-step prediction scenarios. For short-term horizons, MBFNet achieves an E_{RMS} of 0.0418 at the 12-step (60-s) prediction, corresponding to a 132% increase relative to single-step prediction. This increase is notably lower than that of LSTM-attention (145%) and 1D-CNN (122%), indicating slower short-term error accumulation for MBFNet. To further assess long-term predictive performance, the prediction horizon is extended to 24 steps (2 min) and 100 steps (8 min). All models exhibit substantial degradation in accuracy with increasing horizons, reflecting the inherent error accumulation in long-term time-series forecasting (Table 2). For the 24-step prediction, MBFNet achieves an E_{RMS} of 0.0985, representing a 136% increase relative to the 12-step result. At the 100-step horizon, the E_{RMS} further increases to 0.4236, corresponding to a 913% increase compared with the 12-step prediction. Despite this degradation, MBFNet maintains superior relative performance, with its 100-step E_{RMS} being 27.8% lower than that of LSTM-attention. The error growth exhibits clear nonlinear characteristics. Specifically, MBFNet's E_{RMS} increased by 136% from 12 to 24 steps, while the increase from 24 to 100 steps reached 330%, indicating accelerated error accumulation at longer horizons. Additionally, MBFNet reduces E_{RMS} by an average of 24.5% relative to LSTM-attention and 55.2% relative to 1D-CNN across four prediction horizons. This behavior is consistent with observations in complex dynamic systems such as PEMFCs, where long-term prediction uncertainty grows nonlinearly due to coupled electrochemical and operational dynamics.

Table 2 Multi-step prediction performances

Model	Number of steps	R^2	E_{RMS}	E_{MA}	E_{MAP} (%)
MBFNet (ours)	5	0.9421	0.0283	0.0197	0.4821
	12	0.9015	0.0418	0.0294	0.7213
	24	0.8126	0.0985	0.0689	1.8234
	100	0.6347	0.4236	0.2965	7.8924
LSTM-attention	5	0.9132	0.0356	0.0253	0.6347
	12	0.8547	0.0541	0.0389	0.9874
	24	0.7238	0.1347	0.0943	2.4526
	100	0.4825	0.5864	0.4105	10.8923
GRU-attention	5	0.8994	0.0398	0.0281	0.7129
	12	0.8326	0.0612	0.0437	1.1245
	24	0.6842	0.1528	0.1070	2.7845
	100	0.4236	0.6527	0.4569	12.4562
LSTM	5	0.8735	0.0462	0.0329	0.8423
	12	0.7914	0.0725	0.0516	1.3248
	24	0.6124	0.1824	0.1277	3.4527
	100	0.3528	0.7345	0.5142	14.2356
GRU	5	0.8617	0.0491	0.0352	0.9015
	12	0.7739	0.0783	0.0564	1.4527
	24	0.5837	0.1986	0.1390	3.8924
	100	0.3125	0.8123	0.5686	15.8924
1D-CNN	5	0.8213	0.0587	0.0418	1.0724
	12	0.7025	0.0942	0.0683	1.8926
	24	0.4923	0.2345	0.1642	4.5623
	100	0.2438	0.9456	0.6619	18.4527

To evaluate the practical implications of these errors, normalized E_{RMS} values were converted back to the original voltage scale. The normalized E_{RMS} of 0.4236 at the 100-step horizon corresponds to an absolute voltage error of about 12.7 V, based on the 30-V operating range (255.0–285.0 V). Although this error appears large, it represents cumulative deviation over an extended prediction horizon. For energy management system (EMS) applications, which typically rely on prediction horizons ranging from seconds to a few minutes, medium term forecasts (up to 2 min) with absolute errors of about 1.3–3.0 V provide sufficient accuracy for proactive control. These results indicate that MBFNet can provide reliable predictive inputs for EMS decision-making under dynamic automotive operating conditions. Detailed voltage trajectory comparisons for different prediction horizons are provided in Section S4 and Figs. S8–S11 of the ESM.

To further examine predictive stability and physical interpretability, the local prediction performance during voltage recovery intervals is analyzed (Table 3). When extending the prediction horizon from 5 to 12 steps, MBFNet exhibits an R^2 decay of about 4%, which is lower than that observed for GRU-attention and 1D-CNN. During voltage recovery events, MBFNet achieves a local R^2 of 0.8281, significantly outperforming that of GRU-attention (0.4489) and 1D-CNN (0.3025). Furthermore, the correlation coefficient between MBFNet predictions and the actual voltage profile during recovery reaches 0.91, compared with 0.67 for GRU-attention and 0.55 for 1D-CNN. Detailed error analysis is provided in Section S4 and Figs. S12

Table 3 Voltage recovery intervals prediction performance

Model	R^2	E_{MAP} (%)	Correlation	Recovery error
MBFNet (ours)	0.8281	0.45	0.91	1.12×
GRU-attention	0.4489	0.76	0.67	1.84×
1D-CNN	0.3025	1.25	0.55	3.02×

and S13 of the ESM. To verify the contribution of individual MBFNet submodules, an ablation study was conducted using a control-variable strategy. The quantitative results are summarized in Table 4. Removing the multiscale decomposition module (O-multiscale) leads to a 31.7% increase in E_{RMS} and a corresponding 20% increase in voltage recovery phase error, indicating the critical role of temporal decomposition in aligning degradation dynamics across time scales. With or without (w/o) NCT configuration, the model exhibits pronounced noise sensitivity, resulting in elevated E_{RMS} across all prediction horizons, confirming the effectiveness of the proposed denoising mechanism. Removing the bidirectional fusion module (O-fusion) degrades local fluctuation prediction by about 22%, yet still outperforming LSTM/GRU-attention baselines, demonstrating the fundamental validity of the projection-based fusion strategy. Overall, ablations highlight that NCT enhances noise robustness, multiscale decomposition ensures temporal alignment, and bidirectional fusion optimizes feature interaction, enabling MBFNet's performance advantages.

Table 4 Comparison of ablation test performance

Model	E_{RMS}	E_{MA}	Size of parameters ($\times10^4$)	Failure mode
MBFNet	0.0180	0.0123	2.4	None
O-NCT	0.0283	0.0197	1.8	Noise sensitivity
O-multiscale	0.0237	0.0162	1.5	Lagging
O-fusion	0.0205	0.0141	2.1	Local fluctuation

To evaluate deployment feasibility under automotive hardware constraints, a lightweight design analysis was performed by varying the number of projection layers (Table 5). Increasing the number of layers improves the E_{RMS} from 0.0192 to 0.0178 but the marginal accuracy gain diminishes beyond two layers, while memory usage and inference latency increase substantially. A two-layer configuration achieves a favorable trade-off, maintaining E_{RMS} improvement within a 15-ms latency budget suitable for edge deployment.

Table 5 Lightweight design boundary analysis

Number of layers	E_{RMS}	Memory (GB)	Inference latency (ms)
1	0.0192	1.4	8.7
2	0.0180	1.8	12.3
3	0.0178	2.3	18.9

The influence of temporal decomposition depth was further analyzed (Table 6). Increasing the number of decomposition layers from 3 to 5 reduces long-term and short-term E_{RMS} by 61% and 54%, respectively, validating the effectiveness of multiscale modeling. Further increasing the depth to 7 yields limited accuracy gains at the cost of increased computational load. For in-vehicle ECU deployment, a moderate depth (M=5) provides a balanced compromise between accuracy and resource consumption.

Table 6 Comparison of different numbers of decomposition layers (M)

M	Long-term E_{RMS}	Short-term E_{RMS}
3	0.0726	0.0390
5	0.0283	0.0180
7	0.0255	0.0159

Detailed visual comparisons for ablation effects, noise suppression, lightweight configurations, and decomposition sensitivity are provided in Section S4 and Figs. S13–S16 of the ESM.

5 Conclusions

In this paper, a multi-scale bidirectional fusion network is proposed to address the challenges of lifetime prediction of 215-channel PEMFC stacks in automotive scenarios. Key conclusions include

1. A first-of-its-kind GHE co-simulation platform for 215-channel stacks is constructed, enabling synchronized monitoring of flow pulsation and electrochemical parameters, which reduces the bias caused by the distortion of working conditions and scale limitations in traditional laboratories.

2. A joint adaptive noise correlation threshold algorithm is proposed to achieve dynamic suppression of multi-physical field noise without prior modeling, effectively minimizing signal distortion while preserving critical degradation features.

3. A decomposition module is constructed to effectively decouple cross-scale features of voltage recovery and aging processes. Experimental validation demonstrates precise alignment with PEMFC physical degradation mechanisms.

4. A bidirectional fusion module is designed, achieving E_{RMS}=0.0180 during critical degradation (285.0→255.0 V) with an 18.6% improvement over

LSTM-attention. Multi-step prediction shows an average of 24.5% lower E_{RMS} than LSTM-attention and 55.2% lower than 1D-CNN.

5. Lightweight design reduces parameters by 36.8%, meeting stringent ECU resource constraints.

Follow-up work will focus on integrating online diagnostic data to enhance generalizability, developing adaptive parameter adjustment for varying road conditions, and validating robustness via commercial vehicle deployment.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62373321), the Natural Science Foundation of Zhejiang Province (No. LMS25F030013), the National Key R&D Program of China (No. 2022YFB2502405), and the Open Project of Key Laboratory of Automotive Electronics Intelligentization of Zhejiang Province (No. JY20240708), China.

Author contributions

Zifei WANG designed the research and drafted the paper. Xiangxian ZHU designed and tested the corresponding algorithms. Daidai CHEN discussed and revised the algorithms. Congxin LI provided the 215-channel data and processed the corresponding data. Zhitao LIU reviewed and edited the paper. Longhua MA tested the methodology. Jili TAO acquired the funding and administered the project. Hongye SU revised and edited the final version.

Conflict of interest

Zifei WANG, Xiangxian ZHU, Congxin LI, Daidai CHEN, Zhitao LIU, Longhua MA, Jili TAO, and Hongye SU declare that they have no conflict of interest.

Declaration on the use of generative AI tools

During the preparation of this work, the authors used ChatGPT to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Benaggoune K, Yue ML, Jemei S, et al., 2022. A data-driven method for multi-step-ahead prediction and long-term prognostics of proton exchange membrane fuel cell. *Applied Energy*, 313:118835. https://doi.org/10.1016/j.apenergy.2022.118835

Bressel M, Hilairet M, Hissel D, et al., 2016. Extended Kalman filter for prognostic of proton exchange membrane fuel cell. *Applied Energy*, 164:220-227. https://doi.org/10.1016/j.apenergy.2015.11.071

Chen K, Laghrouche S, Djerdir A, 2019. Degradation model of proton exchange membrane fuel cell based on a novel hybrid method. *Applied Energy*, 252:113439. https://doi.org/10.1016/j.apenergy.2019.113439

Chen K, Badji A, Laghrouche S, et al., 2022. Polymer electrolyte membrane fuel cells degradation prediction using multi-kernel relevance vector regression and whale optimization algorithm. *Applied Energy*, 318:119099. https://doi.org/10.1016/j.apenergy.2022.119099

Cheng YJ, Zerhouni N, Lu C, 2018. A hybrid remaining useful life prognostic method for proton exchange membrane fuel cell. *International Journal of Hydrogen Energy*, 43(27): 12314-12327. https://doi.org/10.1016/j.ijhydene.2018.04.160

He K, Liu ZY, Sun YN, et al., 2022. Degradation prediction of proton exchange membrane fuel cell using auto-encoder based health indicator and long short-term memory network. *International Journal of Hydrogen Energy*, 47(82): 35055-35067. https://doi.org/10.1016/j.ijhydene.2022.08.092

He WB, Liu T, Ming WY, et al., 2024. Progress in prediction of remaining useful life of hydrogen fuel cells based on deep learning. *Renewable and Sustainable Energy Reviews*, 192:114193. https://doi.org/10.1016/j.rser.2023.114193

Jouin M, Gouriveau R, Hissel D, et al., 2014. Prognostics of PEM fuel cell in a particle filtering framework. *International Journal of Hydrogen Energy*, 39(1):481-494. https://doi.org/10.1016/j.ijhydene.2013.10.054

Li CZ, Lin W, Wu HY, et al., 2023. Performance degradation decomposition-ensemble prediction of PEMFC using CEEMDAN and dual data-driven model. *Renewable Energy*, 215:118913. https://doi.org/10.1016/j.renene.2023.118913

Liu H, Chen J, Hissel D, et al., 2019. A multi-scale hybrid degradation index for proton exchange membrane fuel cells. *Journal of Power Sources*, 437:226916. https://doi.org/10.1016/j.jpowsour.2019.226916

Liu MY, Wang C, Zhang JB, et al., 2014. Diagnosis of membrane electrode assembly degradation with drive cycle test technique. *International Journal of Hydrogen Energy*, 39(26):14370-14375. https://doi.org/10.1016/j.ijhydene.2014.02.161

Liu ZY, Mao L, Hu ZY, et al., 2022. A novel densely connected neural network for proton exchange membrane fuel cell fault diagnosis. *International Journal of Hydrogen Energy*, 47(94):40041-40053. https://doi.org/10.1016/j.ijhydene.2022.09.158

Ma R, Yang T, Breaz E, et al., 2018. Data-driven proton exchange membrane fuel cell degradation predication through deep learning method. *Applied Energy*, 231:102-115. https://doi.org/10.1016/j.apenergy.2018.09.111

Ma R, Chai XY, Geng RX, et al., 2023. Recent progress and challenges of multi-stack fuel cell systems: fault detection and reconfiguration, energy management strategies,

and applications. *Energy Conversion and Management*, 285:117015. https://doi.org/10.1016/j.enconman.2023.117015

Ren P, Pei PC, Li YH, et al., 2020. Degradation mechanisms of proton exchange membrane fuel cell under typical automotive operating conditions. *Progress in Energy and Combustion Science*, 80:100859. https://doi.org/10.1016/j.pecs.2020.100859

Sahajpal K, Rana KPS, Kumar V, 2023. Accurate long-term prognostics of proton exchange membrane fuel cells using recurrent and convolutional neural networks. *International Journal of Hydrogen Energy*, 48(78):30532-30555. https://doi.org/10.1016/j.ijhydene.2023.04.143

Shin D, Yoo S, 2023. Diagnostic method for PEM fuel cell states using probability distribution-based loss component analysis for voltage loss decomposition. *Applied Energy*, 330:120340. https://doi.org/10.1016/j.apenergy.2022.120340

Sun W, Yan R, Jin R, et al., 2024. LiteFormer: a lightweight and efficient Transformer for rotating machine fault diagnosis. *IEEE Transactions on Reliability*, 73(2):1258-1269. https://doi.org/10.1109/TR.2023.3322860

Tian L, Gao Y, Yang HY, et al., 2025. Multi-scenario long-term degradation prediction of PEMFC based on generative inference informer model. *Applied Energy*, 377:124398. https://doi.org/10.1016/j.apenergy.2024.124398

Tian Z, Wang JH, Al-Durra A, et al., 2023. A novel aging prediction method of fuel cell based on empirical mode decomposition and complexity threshold quantitative criterion. *Journal of Power Sources*, 574:233120. https://doi.org/10.1016/j.jpowsour.2023.233120

Wang C, Dou MF, Li ZL, et al., 2023. Data-driven prognostics based on time-frequency analysis and symbolic recurrent neural network for fuel cells under dynamic load. *Reliability Engineering & System Safety*, 233:109123. https://doi.org/10.1016/j.ress.2023.109123

Wang PH, Liu H, Chen J, et al., 2021. A novel degradation model of proton exchange membrane fuel cells for state of health estimation and prognostics. *International Journal of Hydrogen Energy*, 46(61):31353-31361. https://doi.org/10.1016/j.ijhydene.2021.07.004

Wilberforce T, Olabi AG, 2021. Proton exchange membrane fuel cell performance prediction using artificial neural network. *International Journal of Hydrogen Energy*, 46(8):6037-6050. https://doi.org/10.1016/j.ijhydene.2020.07.263

Wilberforce T, Alaswad A, Garcia–Perez A, et al., 2023. Remaining useful life prediction for proton exchange membrane fuel cells using combined convolutional neural network and recurrent neural network. *International Journal of Hydrogen Energy*, 48(1):291-303. https://doi.org/10.1016/j.ijhydene.2022.09.207

Wu YM, Breaz E, Gao F, et al., 2016. A modified relevance vector machine for PEM fuel-cell stack aging prediction. *IEEE Transactions on Industry Applications*, 52(3):2573-2581. https://doi.org/10.1109/tia.2016.2524402

Xia ZT, Wang YN, Ma LH, et al., 2022. A hybrid prognostic method for proton-exchange-membrane fuel cell with decomposition forecasting framework based on AEKF and LSTM. *Sensors*, 23(1):166. https://doi.org/10.3390/s23010166

Yi FY, Shu X, Zhou JM, et al., 2025. Remaining useful life prediction of PEMFC based on matrix long short-term memory. *International Journal of Hydrogen Energy*, 111:228-237. https://doi.org/10.1016/j.ijhydene.2025.02.302

Zhou DM, Gao F, Breaz E, et al., 2017. Degradation prediction of PEM fuel cell using a moving window based hybrid prognostic approach. *Energy*, 138:1175-1186. https://doi.org/10.1016/j.energy.2017.07.096

Zhou SW, Shearing PR, Brett DJL, et al., 2022. Machine learning as an online diagnostic tool for proton exchange membrane fuel cells. *Current Opinion in Electrochemistry*, 31:100867. https://doi.org/10.1016/j.coelec.2021.100867

Zuo J, Lv H, Zhou DM, et al., 2021. Deep learning based prognostic framework towards proton exchange membrane fuel cell for automotive application. *Applied Energy*, 281:115937. https://doi.org/10.1016/j.apenergy.2020.115937

Electronic supplementary materials

Sections S1–S5, Figs. S1–S17, Tables S1–S3