

Data-driven prognostic algorithm based on Bi-LSTM transformer model for proton exchange membrane fuel cell

L.M. Perez ^{a,b},*, Mohsen Kandidayeni ^b, Alexandre Ravey ^a, Javier Solano ^c, Samir Jemei ^a,
Loïc Boulon ^b

^a Université Marie et Louis Pasteur, UTBM, CNRS, institut FEMTO-ST (UMR 6174), FCLAB, F-90000, Belfort, France

^b Hydrogen Research Institute, Electrical and Computer Engineering Department, Université Du Québec a Trois-Rivières, Québec, G8Z 4M3, Canada

^c European Institute for Energy Research (EIFER), Emmy-Noether-Str. 11, Karlsruhe, 76131, Germany

ARTICLE INFO

Keywords:

Proton Exchange Membrane Fuel Cell
Deep learning prognostics
Time series forecasting
Hybrid neural network architecture
Degradation modelling

ABSTRACT

The complex interaction of internal physical phenomena makes predicting the degradation of Proton Exchange Membrane Fuel Cells (PEMFCs) under dynamic operating conditions a critical but persistent challenge. To address this, the present work develops a hybrid model based on Bidirectional Long Short-Term Memory (Bi-LSTM) layers with a multi-head self-attention mechanism to capture degradation dependencies across both short and long time-frames. When benchmarked against state-of-the-art architectures, including attention-based LSTM variants and the Temporal Fusion Transformer (TFT), the model achieves a reduction of up to 52.7% in Root Mean Square Error (RMSE). The model's accuracy is confirmed across single-step, multi-step, and long-term forecasting scenarios, where it consistently maintains high predictive accuracy ($R^2 \geq 0.998$) for prediction horizons of up to 175 steps ahead. Furthermore, this high performance is achieved with notable data efficiency, requiring a training dataset of 300 h. These combined strengths in accuracy, long-term reliability, and data efficiency make the proposed Bi-LSTM Transformer an effective tool for predictive health management under dynamic load cycles.

1. Introduction

Driven by the escalating impacts of climate change and the significant contribution of the energy sector to greenhouse gas (GHG) emissions, there is a critical imperative to decarbonize global energy systems and mitigate resource depletion [1]. In this context, Microgrids (MGs) and Distributed Renewable Energy Sources (DRES) have emerged as pivotal systems for achieving localized energy reliability and flexibility [2]. However, the integration of these renewable sources into the grid is complex due to the inherent variability and intermittency characterizing generation profiles such as wind and solar power [3]. To mitigate these fluctuations, hydrogen serves as a robust energy vector enabling temporal energy shifting via electrolysis. Complementing this storage capability, Fuel Cells (FCs) reconvert the stored hydrogen into electricity to function as dispatchable generators, thereby stabilizing renewable intermittency [4].

Therefore, a critical application of hydrogen systems represents a key solution for decarbonizing challenging sectors such as transportation [5]. By replacing conventional fossil fuel engines with low-emission alternatives, substantial reductions in emissions can be

achieved across heavy-duty transport, maritime shipping, aviation, and stationary power generation [6]. Among the available technologies, PEMFCs are a compelling solution due to their high energy density, rapid refuelling capability, and absence of harmful emissions [7]. In particular, their suitability for vehicular applications arises from two key advantages. First, their solid proton exchange membrane provides mechanical robustness against vibrations. Second, they operate efficiently at low temperatures, enabling rapid start-ups from ambient conditions [8].

Despite these advantages, ensuring long-term durability remains a critical challenge for PEMFCs, especially when compared to the lifespan of mature technologies like internal combustion engines [9]. This durability is affected by a combination of factors, including material ageing, component design, and a range of operational stressors inherent to vehicular use [10]. The high initial cost of these systems makes predicting the resulting degradation essential to protect the investment and ensure the economic viability of the technology [11].

Prognostic and Health Management (PHM) methodologies offer a promising approach to meet these durability challenges by evaluating

* Corresponding author at: Hydrogen Research Institute, Electrical and Computer Engineering Department, Université Du Québec a Trois-Rivières, Québec, G8Z 4M3, Canada.

E-mail address: Luis.Miguel.Perez.Archila@uqtr.ca (L.M. Perez).

<https://doi.org/10.1016/j.renene.2026.126453>

Received 4 September 2025; Received in revised form 11 September 2026; Accepted 13 September 2026

Available online 18 September 2026

0960-1481/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

the system's current health and predicting its future performance [12]. Such prognostics inform maintenance strategies to improve reliability and extend the operational lifetime [13]. However, creating accurate predictive models is challenging. Their development requires a deep understanding of the complex interplay between electrochemistry, thermodynamics, and fluid dynamics, as well as degradation phenomena that occur over various timescales [14]. Consequently, existing prognostic approaches are broadly categorized into model-based and data-driven [15].

Model-based prognostic methods rely on a detailed understanding of the internal degradation mechanisms affecting various components of PEMFCs. For instance, Koltsiva et al. developed a single-cell mathematical model that captures platinum Electrochemical Surface Area (ECSA) degradation by incorporating processes such as nanoparticle dissolution, migration, and coalescence [16]. Although this model offers a thorough theoretical framework, it still requires extensive experimental validation, which can be resource-intensive. Other studies, such as Zhang et al. [17], have adopted simplified semi-empirical models combined with filtering techniques like the Multi-Point Square-Root Central Difference Kalman Filter (MP-SRCDKF). Nonetheless, their applicability may be limited by assumptions such as constant ageing behaviour over fixed time intervals, which do not always align with real-world operating conditions. Similarly, a constrained version of the Extended Kalman Filter (EKF) is employed in [18]. However, like many model-based approaches, its effectiveness is often restricted to specific operating conditions, which reduces its accuracy, particularly when dealing with complex temporal dependencies.

To overcome the limitations of physics-based models in capturing the complex and non-linear degradation of PEMFCs, researchers have turned to data-based prognostic methods. Many of these methods rely on Artificial Neural Networks (ANN) to automatically learn patterns and trends from experimental data [19]. Initial approaches explored various architectures; for instance, the Adaptive Neuro-Fuzzy Inference System (ANFIS) was used for short-term prognostics [20], while Convolutional Neural Networks (CNN) have been applied for multi-step predictions [21]. For time-series prediction, however, Recurrent Neural Networks (RNNs) became the primary tool due to their inherent ability to process sequential data [22]. Despite their success, traditional RNNs often face challenges, such as the vanishing gradient problem, which complicates the training process. To address this and other limitations, a trend towards hybridization emerged.

This move towards combining multiple architectures is a direct response to the challenge of capturing the multi-scale temporal nature of PEMFC degradation, as will be discussed further in Section 2. Examples include integrating Gated Recurrent Units (GRU) with CNNs [23] or combining LSTM with Support Vector Machines [24]. Beyond simple hybridization, a more advanced approach to mitigate RNN limitations involves the integration of attention mechanisms inspired by Transformer architectures. The Transformer is a deep neural network architecture that replaces the sequential processing of RNNs with a self-attention mechanism. The self-attention mechanism calculates importance scores between all elements, allowing the model to process the entire input sequence simultaneously [25]. This method provides a superior capacity for capturing long-term dependencies and global context, overcoming a key limitation of traditional RNNs [26]. For example, Zuo et al. integrated LSTM and GRU networks with attention modules to predict PEMFC degradation [27]. In another study, LSTM was paired with a self-attention mechanism for single-step-ahead forecasting [28]. However, that approach relied on future known covariates, which limits its applicability to vehicular prognostics, where such information is unavailable. A further development involved combining LSTM with a Dual-stage Attention mechanism (DA-LSTM) to support both multi-step-ahead and long-term prediction tasks [29]. Nonetheless, these models are still constrained by the unidirectional nature of LSTMs, which limits their ability to capture full temporal dependencies. This limitation is problematic in prognostics, as a point's

significance depends on the entire sequence, including backward and forward data points in the sequence. As a result, a fundamental challenge remains: effectively modelling the multi-scale temporal nature of PEMFC degradation, which involves disentangling short-term transient responses from long-term health decline.

Motivated by these limitations, this work proposes a specialized hybrid model that integrates Bi-LSTM layers with a Transformer-inspired self-attention mechanism for improved PEMFC prognostics. The proposed prognostic model introduces a specific architectural adaptation to the standard Transformer framework by replacing sinusoidal positional encoding with Bi-LSTM layers. In the original Transformer architecture, positional encoding is required to provide the model with information about the order of the sequence. Our approach, however, leverages the inherent sequential processing capabilities of Bi-LSTM to create richer, context-aware representations of the input data that implicitly contain this temporal ordering. This adaptation allows the self-attention mechanism to more effectively identify complex dependencies across the entire sequence, thereby enhancing the model's ability to capture both short-term local dynamics and long-term global trends. The primary objective is to develop a robust model capable of accurately forecasting voltage behaviour under diverse operating conditions. To rigorously validate our approach, the proposed model is benchmarked against a curated set of state-of-the-art models from the literature. These benchmarks were specifically chosen to represent key architectural families, including established recurrent baselines like LSTM and GRU, their attention-augmented variants, a modern pure-attention architecture such as the TFT, and other advanced hybrids like the DA-LSTM. All models were compared on the same public dataset, using identical data preprocessing methods and performance metrics, ensuring a comprehensive and fair comparison. The contributions of this study are threefold:

- The implementation of a hybrid neural network architecture that effectively combines the sequential learning capabilities of Bi-LSTM with the global context provided by self-attention mechanisms.
- A comprehensive evaluation of the model's performance through single-step-ahead, multi-step-ahead, and long-term forecasting scenarios.
- Evidence of enhanced predictive accuracy and reliable performance on the tested dynamic load profile, outperforming existing state-of-the-art methods.

The rest of the paper is structured as follows: Section 2 introduces the PEMFC prognostics problem, describing the most significant challenges in predictive modelling. Section 3 describes the experimental dataset, data pre-processing methods, performance metrics, and the proposed model architecture. Section 4 details the results of the prognostic framework along with a comparative analysis across different forecasting strategies. Finally, Section 5 presents the conclusions and outlines directions for future research.

2. The challenge of PEMFC prognostics

The development of an effective prognostic model for PEMFCs requires a deep understanding of the underlying physical phenomena that govern their degradation. This section details these foundational challenges, establishing the context for the data-driven approaches discussed in this work.

2.1. Degradation mechanisms and prognostic challenges

The durability of a PEMFC is not governed by a single, simple process, but by a complex interplay of degradation mechanisms that affect different components over time. These mechanisms are highly sensitive to operating conditions, particularly the dynamic load profiles, start-stop cycles, and humidity levels common in vehicular applications [10].

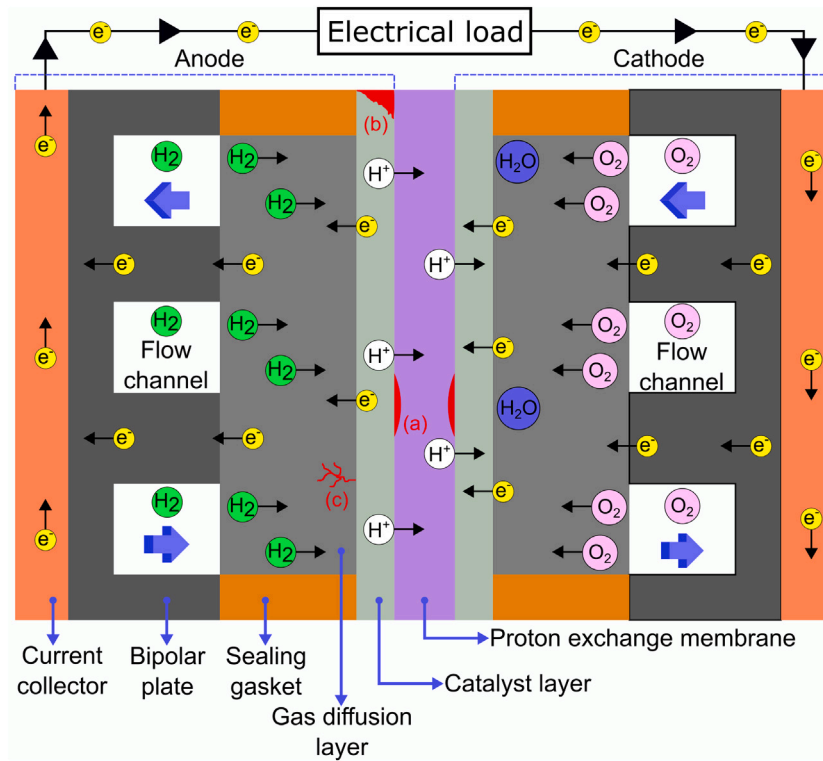


Fig. 1. Operating principle of a PEMFC, illustrating the main components of the MEA. The red markers illustrate some examples of degradation: (a) chemical and mechanical stress on the membrane, (b) platinum dissolution and corrosion in the catalyst layer, and (c) loss of hydrophobicity in the Gas Diffusion Layer (GDL).

The degradation primarily originates in the Membrane Electrode Assembly (MEA), which is composed of the membrane, the catalyst layers, and the Gas Diffusion Layers (GDLs), as illustrated in Fig. 1 [30].

The proton exchange membrane, for instance, is susceptible to both chemical and mechanical stress. Chemical attacks from peroxide radicals (H_2O_2) can thin the polymer structure and create pinholes, while thermal and humidity cycles induce mechanical fatigue, leading to cracks and reactant gas crossover [14]. Simultaneously, the catalyst layers suffer from a loss of Electrochemical Surface Area (ECSA), a critical factor for the PEMFC's efficiency [31]. This loss occurs through multiple pathways, including the dissolution and migration of platinum nanoparticles and the corrosion of the carbon support material, which causes catalyst particles to detach and become inactive [16]. Completing the picture, the Gas Diffusion Layers (GDLs) can lose their hydrophobicity over time, which impairs water management and leads to electrode flooding, while mechanical compression can reduce their porosity, hindering effective gas distribution [10].

These interconnected physical degradation phenomena lead to a gradual and complex loss of performance, presenting several significant challenges for predictive models [32]. Firstly, the rate of degradation is not constant. It is strongly modulated by dynamic operating conditions that create time-varying stress on the components [33]. This variability is difficult to capture with simple empirical models [34]. Secondly, the degradation path itself is often non-linear and non-monotonic. The degradation path can alternate between accelerated decay, slow degradation, and even moments of partial performance recovery [35]. This complexity can easily mislead prognostic algorithms. Finally, as shown in Fig. 2, the resulting operational data exhibits multi-scale temporal features. This means the signal contains rapid, transient voltage responses to load changes (on the order of seconds or minutes) superimposed on the slow, underlying degradation trend (on the order of hundreds of hours). A robust prognostic model must therefore be capable of disentangling these short-term dynamics from the long-term health decline to make accurate forecasts.

2.2. Forecasting approach

The primary goal of a prognostic model is to forecast the evolution of the PEMFC's performance. State of Health (SOH) is a measure of the present performance condition of a PEMFC [13]. The output voltage is the most direct and non-invasive performance indicator. This is because it is constantly monitored during normal operation and encapsulates the final effect of all internal electrochemical and degradation processes [15]. Therefore, a model that accurately predicts future voltage behaviour is implicitly forecasting the system's SOH. Under dynamic load profiles, this voltage fluctuates constantly. As a result, the SOH is not reflected in a single static value but rather in the dynamic relationship between the demanded current and the resulting voltage over time [36].

However, the unknown nature of future load profiles in many vehicular applications poses a significant challenge for forecasting [37]. To address this, the approach taken in this work uses a neural network to directly map the historical voltage sequence to the future voltage sequence, a method known as an autoregressive approach. By learning from past voltage data alone, the model implicitly learns to distinguish between the effects of the load and the underlying degradation trend, demonstrating its capacity to capture the system's health state and degradation trajectory. Crucially, this autoregressive approach avoids the need for known future inputs, such as the current profile. This stands in contrast to alternative methodologies that require future covariates to make a prediction, a fundamental difference that is illustrated conceptually in Fig. 3.

Although the proposed model is fundamentally data-driven and does not explicitly modify the electrochemical governing equations, it operates on the premise that physical degradation mechanisms imprint a distinct signature on the voltage dynamics. As discussed in Section 2.1, phenomena such as catalyst degradation (ECSA loss) and membrane ageing induce concrete alterations in internal resistance and reaction kinetics. These physical modifications manifest as observable

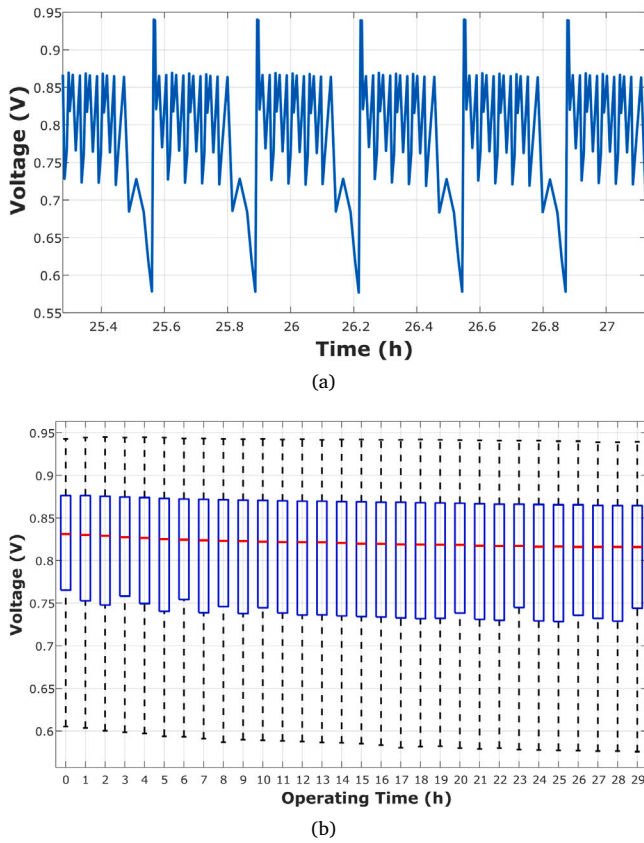


Fig. 2. PEMFC voltage data exhibiting multi-scale temporal characteristics, showing short-term dynamics (top) versus the long-term degradation trend visualized with hourly box-plots (bottom).

deviations in the voltage response, characterized by increased voltage drops during load transients or high current regimes. By processing the historical voltage sequence, the model learns to distinguish these multi-scale temporal patterns, associating short-term transient responses with immediate load conditions while linking long-term drift to the accumulation of irreversible physical damage.

3. Materials and methods

This section is divided into three subsections. First, we describe the experimental dataset on which the performance of the proposed forecasting algorithm has been tested. Second, we present the data pre-processing used, and the performance metrics adopted to measure the model's accuracy. Finally, we explain the model architecture along with the theoretical basis of the key components.

3.1. Experimental dataset

The experimental data used in this study is an open-access dataset reported in [38]. It contains long-term durability test results performed on a commercial single-cell PEMFC, manufactured by Wuhan WUT New Energy Co. The durability tests were conducted using a Greenlight fuel cell test station (G20, P17-2390, Burnaby, B.C., Canada), as illustrated in the schematic in Fig. 4. During testing, the PEMFC operated under a Fuel Cell Dynamic Load Test Cycle (FC-DLC), designed based on the New European Driving Cycle (ECE R15), as illustrated in Fig. 5. This dynamic load cycle was repeated 3076 times over the entire dataset, resulting in approximately 1008 h of operational data recorded at a sampling frequency of 1 Hz.

Each FC-DLC lasts 1180 s and consists of dynamic current variations, during which the current changes through 35 discrete steps and assumes one of nine values: 0, 1.76, 4.43, 9.48, 10.38, 14.82, 20.70, 29.59, and 35.53 A. The dataset includes measurements of 17 variables, such as time, current, and voltage, along with 12 additional variables related to the fuel cell's operating conditions. These additional parameters include pressures, temperatures, and volumetric flow rates measured at the electrodes, and their evolution throughout the experiment is shown in Fig. 6.

As observed in Fig. 6, particularly in subplots (e) through (l), the different temperature variables are maintained within ranges close to the nominal operating conditions for PEMFCs. This thermal stability indicates that the degradation observed during the experiment is not attributed to water phase changes, freezing, or extreme temperature excursions. Therefore, the scope of the proposed prognostic model is defined for PEMFCs operating within nominal thermal ranges, and does not encompass degradation caused by extreme thermal changes.

Among the measured variables, voltage is the most commonly used health index in the literature. This choice is mainly justified by the fact that the voltage across the stack terminals is directly influenced by operating conditions and closely reflects the degradation processes occurring within the PEMFC [17]. Moreover, relying solely on voltage transforms the forecasting task into an auto-regressive problem, enabling multi-step ahead predictions without the need for future values of external variables. Given these advantages, voltage has been selected in this study as the key health index to assess performance degradation over time.

3.2. Data pre-processing

Due to the large volume of data (1008 h sampled once per second), managing computational resources and training the neural network present significant challenges. In particular, high noise can generate poor predictions, especially in predicting long-term trend degradation. As a mitigation measure, a resampling procedure was carried out. In each cycle, 35 representative values were selected to capture the 35 transitions available in the FC-DLC. To reduce noise from transient voltage oscillations, the fifth-to-last data point within each constant current interval was chosen. This specific sampling point was selected to capture the quasi-steady state voltage, effectively filtering out short-term electrical transients. This criterion is also consistent with previous prognostic studies. Zuo et al. [27] first extracted one representative point per current step of the FC-DLC to build the voltage health indicator on this same dataset. Li et al. [28] applied the same per-step extraction to the operating covariates feeding their temporal fusion transformer. Yang et al. [39] used an equivalent quasi-steady-state sampling under static, quasi-dynamic and dynamic conditions. Adopting it here keeps the present results directly comparable with these benchmarks. Overall, this procedure, applied over 3076 FC-DLC cycles, resulted in approximately 107,600 data points. The resulting PEMFC voltage after resampling is shown in Fig. 7.

The second step of the data pre-processing is the division of the dataset into training (X_{train}) and test (X_{test}) sets. In order to enhance the training performance of the neural network, the data are normalized by a standard scaler so that all of the features follow a normal distribution with a mean of 0 and a standard deviation of 1, as is defined as follows:

$$x_i = \frac{x_i - \bar{x}_{\text{train}}}{\sigma_{\text{train}}}, \quad x_i \in X_{\text{train}} \cup X_{\text{test}} \quad (1)$$

$$\bar{x}_{\text{Train}} = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} x_i, \quad x_i \in X_{\text{Train}} \quad (2)$$

$$\sigma_{\text{Train}} = \sqrt{\frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} (x_i - \bar{x}_{\text{Train}})^2}, \quad x_i \in X_{\text{Train}} \quad (3)$$

Here, x_i are the input data. $\bar{x}_i$ are the normalized data. N_{train} denotes the number of data points in the training set. $\bar{x}_{\text{Train}}$ is the mean,

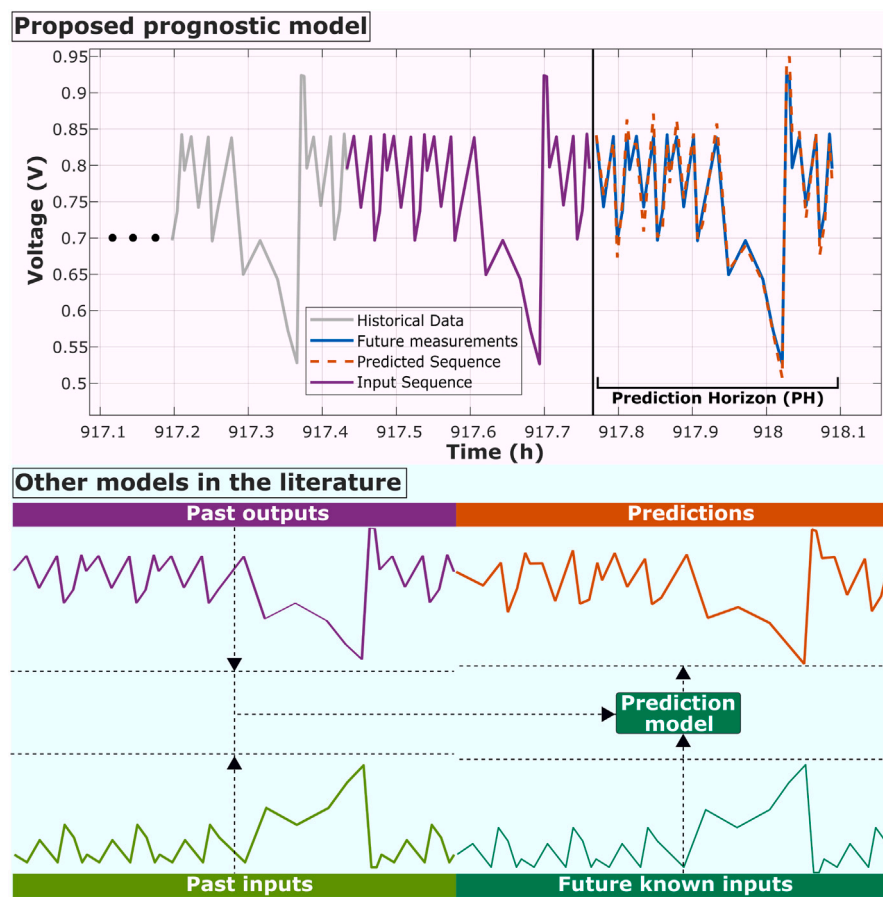


Fig. 3. Comparison of forecasting methodologies: the proposed approach using only historical data (top), versus an alternative approach requiring future known inputs (bottom).

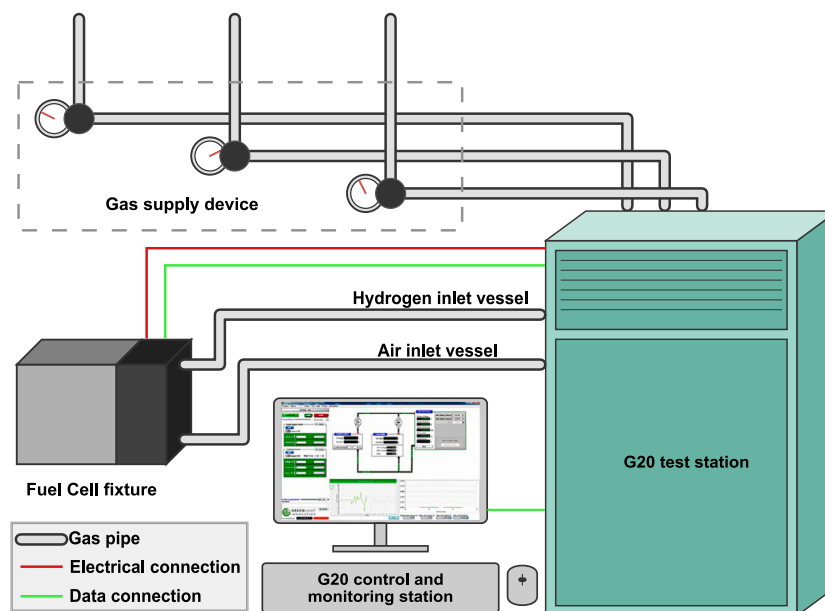


Fig. 4. Schematic of the experimental test bench used for data acquisition [27].

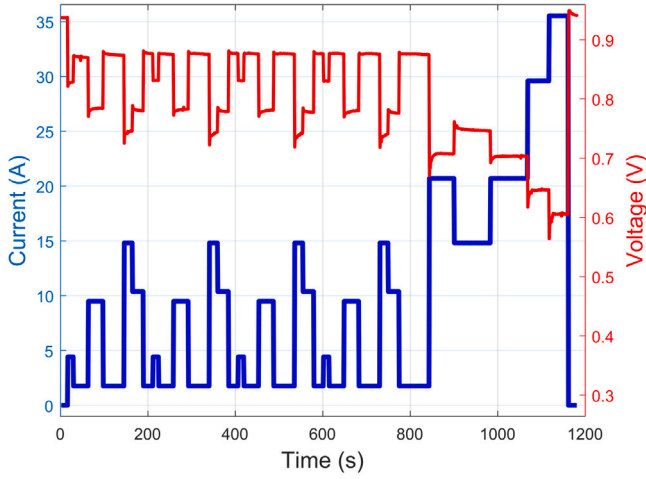


Fig. 5. Dynamic voltage and current to one FC-DLC.

and σ_{Train} is the standard deviation of the training set. Standardization is performed on the training and test sets. Finally, each dataset is organized into input sequences defined by a fixed time window (S) and target output sequences of a given length (T).

3.3. Performance evaluation metrics

To evaluate the performance of the proposed model, four commonly used metrics in time series forecasting are considered: Mean Squared Error (MSE), RMSE, MAE, and the R^2 . The mathematical formulations for each of these metrics are presented below.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

Here, y_i represents the actual observed values, while $\hat{y}_i$ corresponds to the predicted values. The total number of samples is denoted by n , and $\bar{y}$ represents the mean of the observed values in the test set. A lower MSE, RMSE, and MAE show a higher predictive accuracy, whereas an R^2 value approaching 1 suggests a stronger correlation between the predicted and actual values.

3.4. Bi-LSTM Transformer model

This section describes the architecture of the proposed Bi-LSTM Transformer model, which incorporates Bi-LSTM layers into the conventional encoder of the classic Transformer framework, along with the attention mechanism. The complete model, designed for time series prediction, is illustrated in Fig. 8.

In this architecture, the Bi-LSTM layers take the place of the conventional positional encoding and feedforward modules, thereby enabling a deeper capture of temporal dependencies. In the following subsections, complete explanations of all layers are presented, such as the operation mechanism of the Bi-LSTM layers and the operation of the multi-head self-attention layer.

Table 1

Parameters for the proposed neural network.

Parameter	Value
Hidden units of 1st Bi-LSTM layer	35
Number of attention heads	4
Dimension of keys and queries	16
Neurons in the 1st dense layer	70
Hidden units of 2nd Bi-LSTM layer	35
Neurons in the 2nd dense layer	T
Optimizer	Adam
Epochs	100
Gradient threshold	1
Dropout factor	0.1

3.4.1. Bi-LSTM

LSTM networks are a special subset of recurrent neural networks designed to avoid the vanishing gradient problem, thereby enabling the efficient handling of long-range dependencies in sequences [40]. In the context of time series forecasting, each LSTM cell operates with two key states: the hidden state (h_t), which is also the output at time step t , and the cell state (C_t), which accumulates information from previous time steps via (C_{t-1}) and (h_{t-1}) from other cells in the same layer. At each time step, the cell state is updated by selectively adding or removing information through a set of gating mechanisms: the forget gate (f_t), the input gate (i_t), the cell candidate (g_t), and the output gate (o_t). The gating mechanisms depend on learnable parameters: the input weights (W) associated with the input layer (x_t), the recurrent weights (R) related to previous hidden states, and the bias b . Their mathematical definitions are as follows:

$$f_t = \sigma(W_f \cdot x_t + R_f \cdot h_{t-1} + b_f) \quad (8)$$

$$i_t = \sigma(W_i \cdot x_t + R_i \cdot h_{t-1} + b_i) \quad (9)$$

$$g_t = \tanh(W_g \cdot x_t + R_g \cdot h_{t-1} + b_g) \quad (10)$$

$$o_t = \sigma(W_o \cdot x_t + R_o \cdot h_{t-1} + b_o) \quad (11)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot g_t \quad (12)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (13)$$

The general structure of the LSTM cell is shown in Fig. 9. Based on the LSTM cell, the Bidirectional Long Short-Term Memory (Bi-LSTM) layer improves upon sequence modelling by simultaneously examining the input data in both forward and backward directions with two parallel LSTM networks. The dual-path mechanism encourages the network to take information from both past and future contexts and thus generates a more comprehensive representation of the input sequence. For this reason, Bi-LSTM layers are especially useful for tasks involving temporal analysis, such as predicting complex time series data.

3.4.2. Self-attention mechanism

Attention mechanisms have become a fundamental component of modern deep learning models, particularly in sequence data modelling. Unlike traditional methods, which present challenges connecting distant elements in a sequence, attention layers allow the model to compare all parts of the sequence directly. This property allows it to efficiently capture global contextual information [41].

Self-attention is a mechanism used to capture relationships between different elements within the same input sequence, helping the model build a more meaningful representation. It works by comparing each

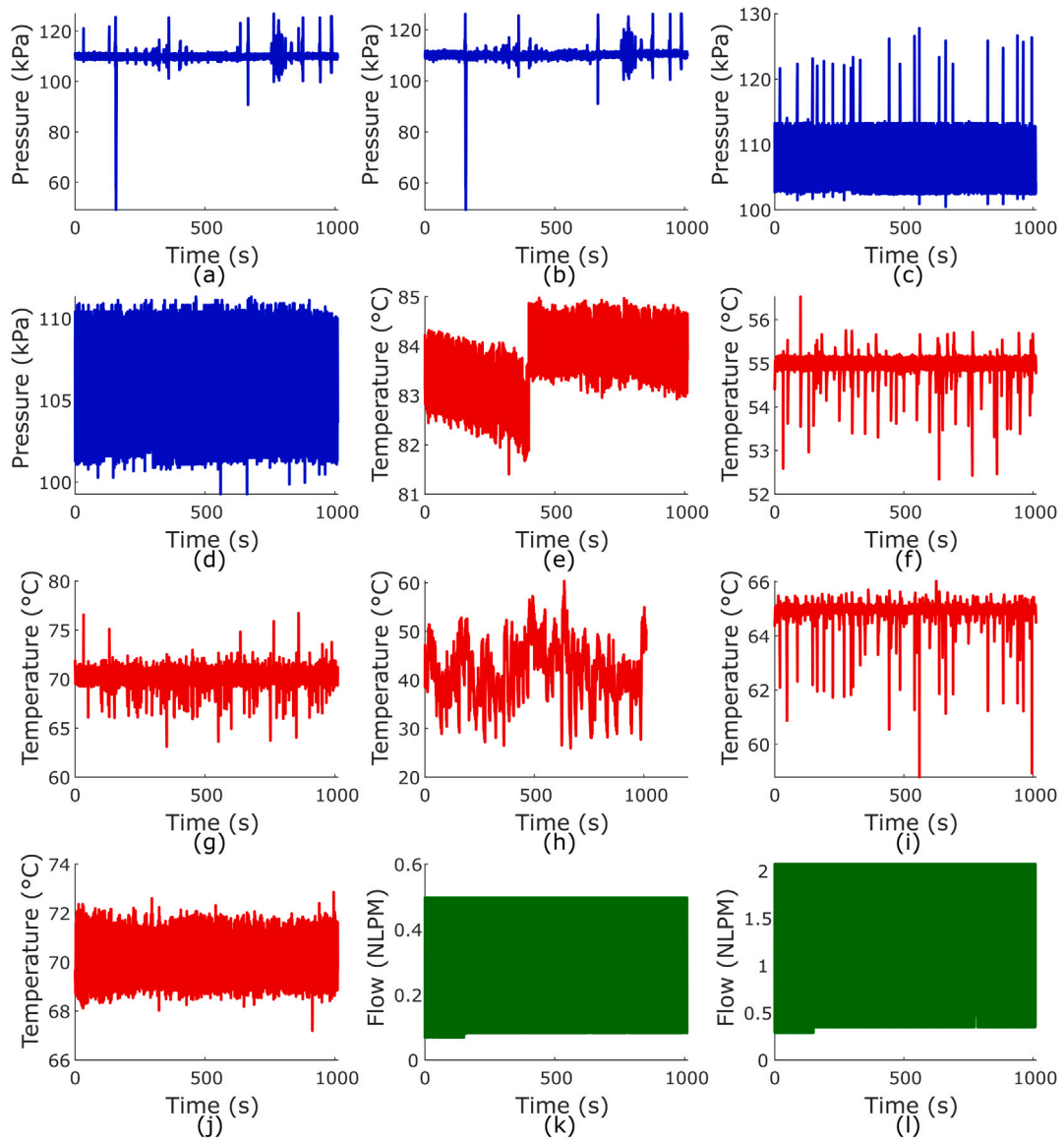


Fig. 6. Operating conditions of the dataset with the FC-DLC. (a) Anode inlet pressure. (b) Anode outlet pressure. (c) Cathode inlet pressure. (d) Cathode outlet pressure. (e) Anode endplate temperature. (f) Anode water dew point temperature. (g) Anode inlet temperature. (h) Anode outlet temperature. (i) Cathode water dew point temperature. (j) Cathode inlet temperature. (k) Total anode flow. (l) Total cathode flow.

element, called a query, with all other elements in the sequence, known as keys, and uses this comparison to calculate attention weights. The weights determine the amount of attention each item must pay to others. The result is a weighted sum of the elements, also called values, by these attention weights [25].

For the scaled dot-product attention mechanism, queries, keys, and values are all vectors that come from the input. The attention weights are calculated by calculating the dot product between the query and each key, scaling it, and applying a softmax function. This assists the model in giving more importance to the most critical parts of the sequence. The mathematical formula for the single-head scaled dot-product attention is given as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}(\lambda QK^T) V \quad (14)$$

where Q , K , and V denote the matrices of queries, keys, and values, respectively, and λ is one scaling factor.

To further enhance the model's ability to capture diverse information, multi-head self-attention is employed. The multi-head self-attention operation for an input X is defined as:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_O, \quad (15)$$

The output of each head is computed as:

$$\text{head}_i = \text{Attention}(XW_i^Q, XW_i^K, XW_i^V). \quad (16)$$

Here, h denotes the number of attention heads; W_Q , W_K , and W_V are learnable projection matrices for queries, keys, and values, respectively; and W_O is a learnable projection matrix for the output.

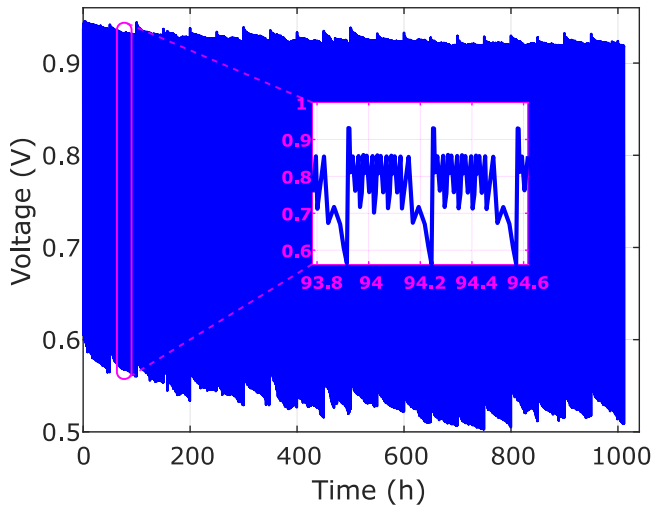


Fig. 7. Voltage over the full dataset with resampling.

This multi-head approach enables the model to simultaneously focus on different parts of the input sequence, thereby capturing a richer set of features and dependencies.

4. Results

The proposed prognostic algorithm was configured through a grid search method over the hyperparameters presented in Table 1. Systematic optimization enabled the identification of the best model configuration for the prediction task.

This section discusses the predictive performance of the previously described data-driven prognostic framework. Three different prediction strategies are evaluated: single-step-ahead, multi-step-ahead, and long-term prognostic forecasting. The primary differences between them are the number of forecast output steps (T) and the prediction horizon (PH). Specifically, when the model predicts only the next time step in the series ($T = 1$), it is referred to as single-step-ahead prediction. In multi-step-ahead forecasting, the predicted sequence length is greater than or equal to the prediction horizon ($L \leq PH$). For long-term prognostics, the length of the predicted sequence is shorter than the prediction horizon ($L > PH$); hence, the prediction is performed iteratively over the horizon. An illustration of these prediction strategies is provided in Fig. 10.

All the simulations carried out in this research were run on MATLAB 2024a on a computer equipped with an i7-6700 processor and 16 GB RAM. The following subsections will discuss the experimental results, first demonstrating the performance of single-step prediction, then analysing multi-step and long-term forecasting.

4.1. Single-step-ahead prediction

For the single-step-ahead experiments, the dataset was split into 60% for training and 40% for testing. In this configuration, the model receives an input sequence consisting of the 5 most recent time steps, a selection guided by the methodology in [28], and the forecasting task involves predicting the very next time step ($T = 1$). In this case, the prediction horizon is equal to the length of the output sequence ($T = PH$), as only the next point in the time series is estimated. Fig. 11 illustrates a representative single-step-ahead prediction obtained at 600 training hours, while Table 2 summarizes the corresponding performance metrics on both the training and test sets.

The performance metrics were employed to compare the predictive performance of the suggested model with that of various state-of-the-art benchmarks from the literature. They are the TFT [28], Mixed-Gated Unit (MIXGU) [39], Attention Mixed-Gated Unit (AT-MIXGU) [39], LSTM [27], LSTM with attention mechanism [27], GRU [27], and GRU with attention mechanism [27]. The results demonstrate that our proposed algorithm achieves improvements in the prediction accuracy, for example: 52.7% reductions in RMSE, 80% in MSE, and 49.2% in MAE. In addition, higher R^2 values were obtained, confirming its superior capabilities in single-step forecasting. The training time of the Bi-LSTM Transformer model with this training window is 10,325 s.

To evaluate the robustness of the proposed algorithm with respect to the size of the training dataset, a series of experiments was carried out using training windows of 300, 400, 500, 600, 700, and 800 h. The development was compared with the DA-LSTM model as presented in [29], and the performance metrics are summarized in Fig. 12.

Across all training durations, the proposed model performs better than the DA-LSTM model. In terms of RMSE, the improvements range from a 68.9% reduction at 300 h to a peak reduction of 80.2% at 700 h. Likewise, MAE values show reductions between 61.8% and 74.9%, implying significantly lower average magnitude of error throughout the testing phase.

It is particularly interesting to notice that the model behaves with high accuracy, even on smaller amounts of training data. For example, with just 300 h of training, the proposed model has an RMSE of 0.0037 and an MAE of 0.0025, while DA-LSTM has 0.01189 and 0.00657, respectively.

In addition, the coefficient of determination (R^2) stays consistently high for the proposed model, always above 0.998, and shows improvements of up to 1.6% compared to the DA-LSTM. These findings not only demonstrate the effectiveness of the suggested architecture in reducing prediction error but also demonstrate its stability for different levels of training data, which is an important feature in prognostic applications when historical data is limited.

4.2. Multi-step-ahead prediction

In this subsection, the performance of the model proposed is analysed for multi-step-ahead forecasting and compared to the results reported in [29]. The test set consists of the last 10,000 resampled data points, while the remaining data are used for the train phase. The input sequence length is fixed at $S = 5$, and the output sequence length ranges from 1 to 5 steps ($T = 1, 2, 3, 4, 5$), where the prediction horizon is equal to the number of steps ahead ($T = PH$). Fig. 13 shows the forecasting results for horizons ranging from 1 to 5 steps.

In the left column, are presented the measured and predicted voltage over the test set. In the right column, the absolute error between the prediction and measured voltage stack over the full dataset. Additionally, Fig. 14 summarizes the corresponding accuracy metrics.

As expected, prediction errors tend to increase with the number of time steps ahead. However, when the output sequence length is set equal to the prediction horizon, for example in the transition from 4-step to 5-step predictions, a slight reduction in error is observed. Furthermore, a comparison with the DA-LSTM results in [29] is performed. For the 1-step-ahead prediction, DA-LSTM reported an MAE of 0.0025 and an R^2 of 0.996, while our model achieved a 64.4% lower MAE and a 0.4% higher R^2 . When the prediction horizon is set to 5 steps, DA-LSTM obtained an MAE of 0.0061 and an R^2 of 0.989, whereas the developed Bi-LSTM Transformer model reduced the MAE by 57.3% and improved R^2 by 1% compared to the literature values. These findings indicate that the model consistently offers greater accuracy at different horizons of prediction and validate the robust stability of the Bi-LSTM Transformer model in multi-step forecasting scenarios.

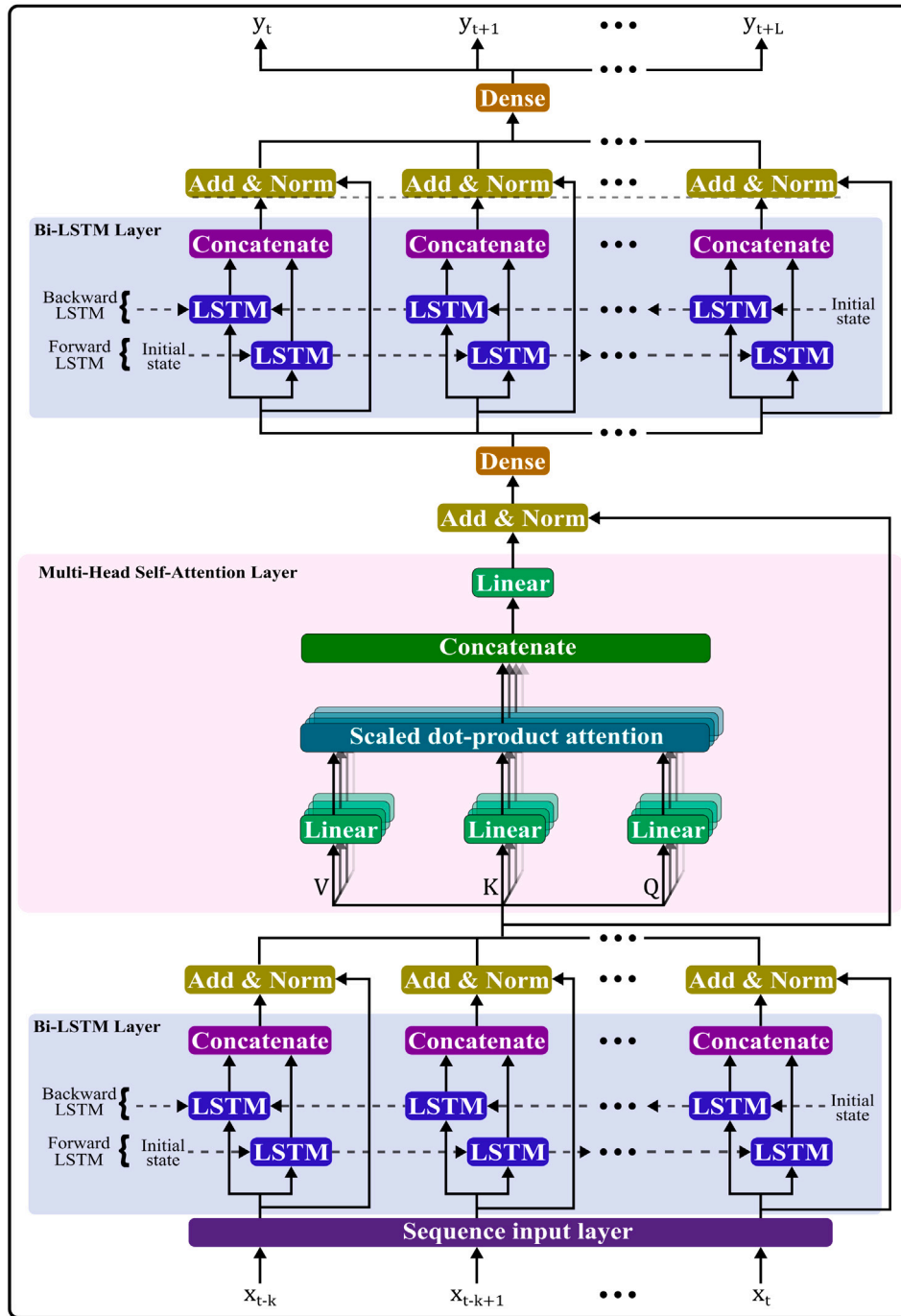


Fig. 8. Bi-LSTM Transformer model.

Table 2
Performance metrics of single-step-ahead predictions.

	Proposed	TFT [28]	MIXGU [39]	AT-MIXGU [39]	LSTM [27]	LSTM attention [27]	GRU [27]	GRU attention [27]
Training set								
RMSE	0.001536	0.003247	0.011622	0.009975	0.011066	0.008542	0.015873	0.013091
MSE	0.000002	0.000010	–	–	0.000122	0.000073	0.000252	0.000171
MAE	0.001064	0.002094	0.007432	0.006245	0.007549	0.006170	0.011751	0.009827
R^2	0.999694	0.998602	0.993130	0.995540	0.984143	0.990592	0.967374	0.977808
Test set								
RMSE	0.002471	0.006721	0.012181	0.011049	0.017637	0.016409	0.018260	0.0115518
MSE	0.000006	0.000045	–	–	0.000311	0.000269	0.000333	0.000241
MAE	0.001681	0.003568	0.007914	0.007014	0.013970	0.013002	0.014368	0.012016
R^2	0.999265	0.991109	0.994540	0.994630	0.962568	0.967600	0.959877	0.971025

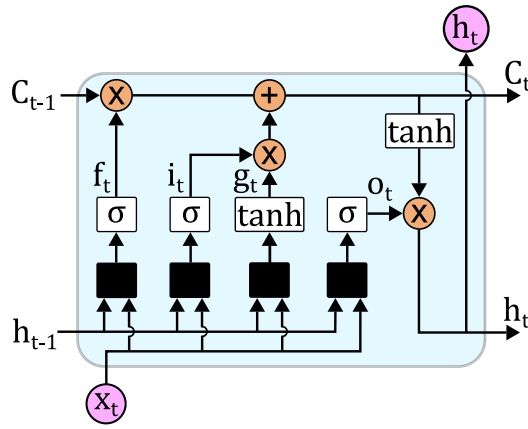


Fig. 9. Structure of a LSTM cell.

4.3. Long-term prognostic

This subsection examines the performance of the proposed model under long-term forecasting conditions, where the prediction horizon is much larger than the output sequence. The training set is fixed to approximately 300 h, and the test set comprises the remaining data. The input sequence length is increased to ($S = 35$), and the output sequence length is set equal to this value ($T = 35$). The window size of $S = 35$ was chosen to correspond with the length of one full dynamic load cycle. The forecasting task considers prediction horizons (PH s) set to 70, 105, 140, and 175, each significantly exceeding the output sequence length.

Fig. 15 presents the results of simulation. Because of the variability and complexity present in the complete dataset, the left-hand column of the figure plots only the voltage for the nominal current condition ($I = 35.6$ A) for visual clarity. The right-hand column shows the absolute error between the predicted voltage and the actual measured values across the test set for each prediction horizon.

During the first 300 h of the forecast interval, the proposed model reflects the overall trend of degradation, even as the forecasting horizon increases. A substantial increase in prediction error takes place between 600 and 800 h. This increase is due to degradation-slope changes and recovery-voltage phenomena that create abrupt positive spikes in the measured signal. These variations become more meaningful as the prediction horizon lengthens, reflecting the growing temporal uncertainty typical of long-term predictions. Despite this, the absolute error is below 10% of the nominal voltage for the entire test set, indicating that the model maintains an acceptable level of accuracy for all the prediction horizons.

To study the ability of the model to capture meaningful physical correlations consistent with PEMFC dynamics, an analysis of the self-attention weights was conducted. Fig. 16 presents a comparative visualization of the global attention maps under two distinct operating regimes defined as Low Current and High Current.

As observed in Fig. 16, the attention mechanism dynamically adapts its focus based on the operating condition. Under the Low Current conditions shown in Fig. 16(a), the attention weights are distributed evenly across the earlier and middle sections of the time window. This pattern suggests a stable behaviour where the model relies on a broader history to characterize the steady-state performance since the system exhibits lower volatility. Conversely, under high current conditions as depicted in Fig. 16(b), the mechanism concentrates heavily on the most recent time steps ($t = 35$) located in the upper region of the graph. This shift indicates a reactive behaviour where the model correctly

Table 3

Performance metrics of the proposed model for the long-term forecasting task across various PHs on the test set.

	PH = 70	PH = 105	PH = 140	PH = 175
Test set				
RMSE	0.0036	0.0045	0.0054	0.0057
MSE	1.29×10^{-5}	2.06×10^{-5}	2.87×10^{-5}	3.27×10^{-5}
MAE	0.0022	0.0027	0.0033	0.0036
R^2	0.9984	0.9975	0.9965	0.9960

identifies that the immediate history is far more predictive of the next voltage drop than the distant past during high-load degradation phases. This observation aligns with the fast dynamics of ohmic losses active at medium and high currents, which are driven by transport phenomena related to membrane water content and proton exchange. However, it is important to explicitly acknowledge that despite the insights provided by attention mechanisms, neural network architectures are fundamentally data-driven black-box models. Consequently, full physical interpretability is limited, as the model does not explicitly resolve the physical equations governing the degradation mechanisms. The numerical results for each prediction horizon are summarized in Table 3.

The performance metrics show that prediction error generally increases as the prediction horizon extends. For example, RMSE rises from 0.0036 at $PH = 70$ to 0.0057 at $PH = 175$, an increase of approximately 58%. Similarly, MAE grows from 0.0022 at $PH = 70$ to 0.0036 at $PH = 175$, an increase of about 64%, while MSE escalates from 1.29×10^{-5} to 3.27×10^{-5} , corresponding to roughly 153%. On the other hand, the R^2 values show only a slight decline, remaining consistently high (from 0.9984 at $PH = 70$ to 0.9960 at $PH = 175$), which underscores the model's capacity to capture the variance in the data even under extended prediction scenarios.

5. Conclusions

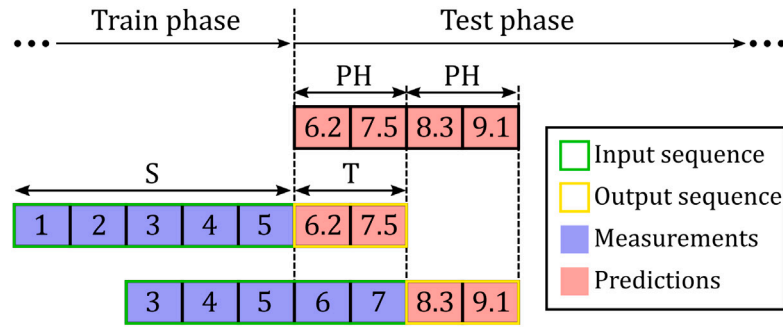
This paper proposed a specialized prognostic framework based on a hybrid Bi-LSTM Transformer model for accurate voltage prediction in PEMFC under dynamic operations. The model was designed to acquire temporal dependencies and dynamic patterns in multivariate time-series data through the combination of the sequential processing capability of bidirectional LSTM layers and contextual understanding provided by self-attention mechanisms.

Three prediction strategies were examined to evaluate the model's robustness and generalization: single-step-ahead, multi-step-ahead, and long-term forecasting. In the single-step-ahead task, the model consistently outperformed state-of-the-art benchmarks from the literature, achieving significant reductions in RMSE (up to 52.7%), MSE (up to 80%), and MAE (up to 49.2%), while maintaining a high R^2 above 0.9997. To evaluate the learning capabilities of the proposed model, different ratios between the train and test sets were employed. The Bi-LSTM Transformer model remained at an R^2 of up to 0.998, even when trained with 300 h of data. These results highlight the model's ability to deliver accurate forecasts in scenarios where historical data is limited.

In the multi-step-ahead forecasting strategy, the proposed model exhibited consistent and reliable performance across increasing prediction horizons. Particularly, for 5-step-ahead predictions, it achieved a 57.7% reduction in MAE and a 1% improvement in R^2 , with all R^2 values remaining above 0.9993. These results confirm the model's robustness and its capacity to deliver accurate forecasts over short and moderate-term horizons.

In the long-term prognostic scenario, the Bi-LSTM Transformer continues to track the underlying voltage trend even though the prediction horizon is considerably larger than the output window. Interestingly, this performance is accomplished using 300 h of training data, speaking to the strength of the model's learning capacity as well as its ability to

Multi-step-ahead prediction



Long-term prognostics

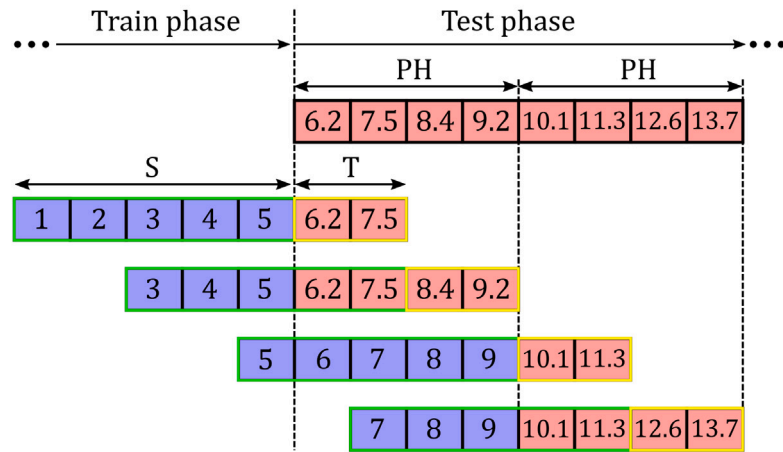


Fig. 10. Demonstration of the prediction strategies: multi-step-ahead and long-term prognostics.

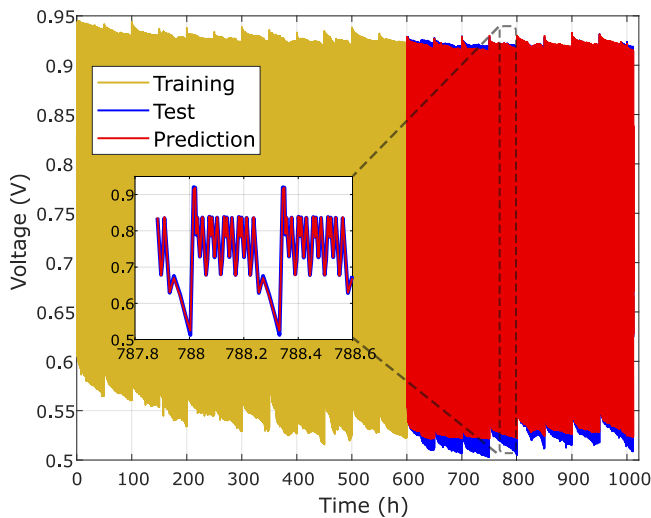


Fig. 11. Single-Step-Ahead prediction at 600 h of train.

model both local and global temporal dependencies through its hybrid architecture. Prediction accuracy is maintained at an acceptable level (absolute error below 10% of the nominal voltage), and all R^2 values are above 0.996, even for the longest horizon considered. Although the model still struggles with abrupt events such as recovery peaks, stability and accuracy are preserved over long time periods. While the findings demonstrate the model's capabilities on this dynamic profile (FC-DLC), future work will consider further validation on datasets with different load characteristics and failure modes to confirm the generalization of the proposed framework.

In summary, the proposed Bi-LSTM Transformer-based approach is a robust and accurate prognostic alternative for PEMFCs under the tested conditions. Its accuracy under different forecast scenarios makes it an effective tool for predictive maintenance and condition monitoring applications in real working conditions. Future work will focus on several key areas to build upon the current model. A primary direction is the integration into health-conscious energy management strategies, where the model's predictions could adjust operating parameters in real-time to minimize degradation and extend the system's lifespan. Another crucial step is exploring real-time implementation, which involves optimizing the model's architecture for computational efficiency to meet the requirements of on-board vehicular systems without significantly compromising accuracy. Furthermore, the model's robustness will be validated by testing it on more diverse datasets, including different driving cycles and data from various fuel cell stacks and sizes.

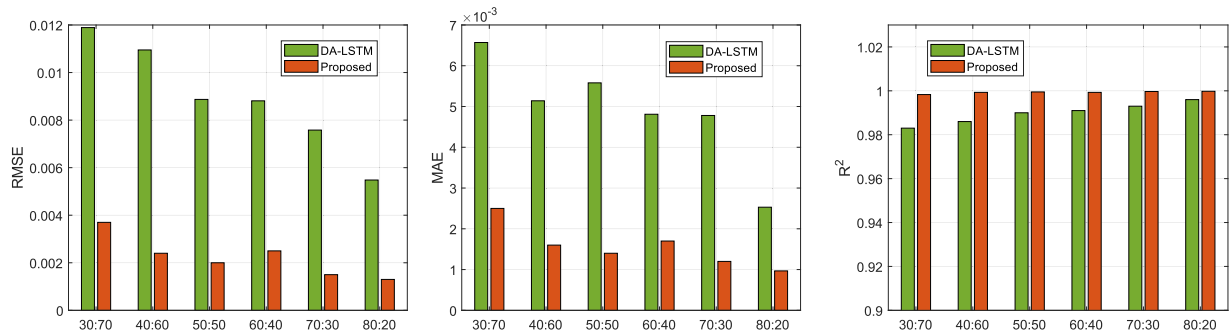


Fig. 12. Accuracy metrics for predictions with different training data windows.

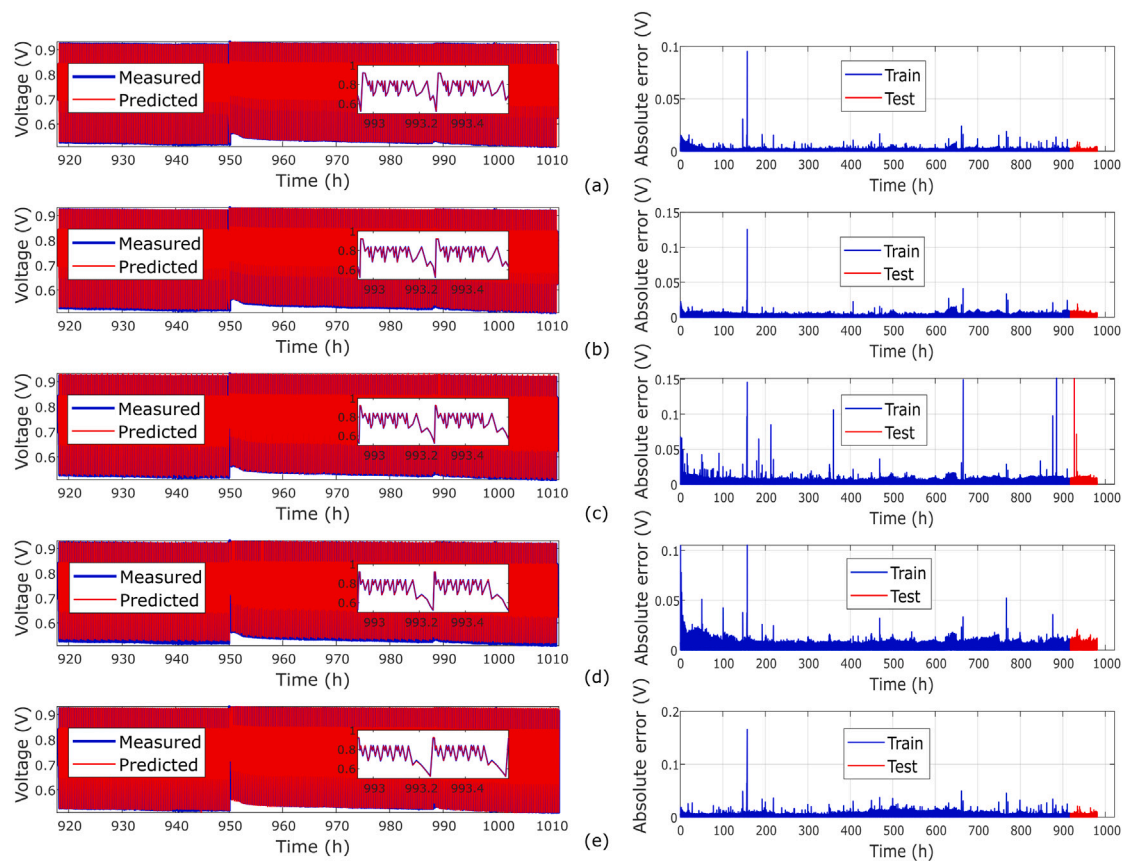


Fig. 13. Predicted versus actual outputs for multi-step-ahead forecasting scenarios: (a) 1-step ahead, (b) 2-step ahead, (c) 3-step ahead, (d) 4-step ahead, and (e) 5-step ahead.

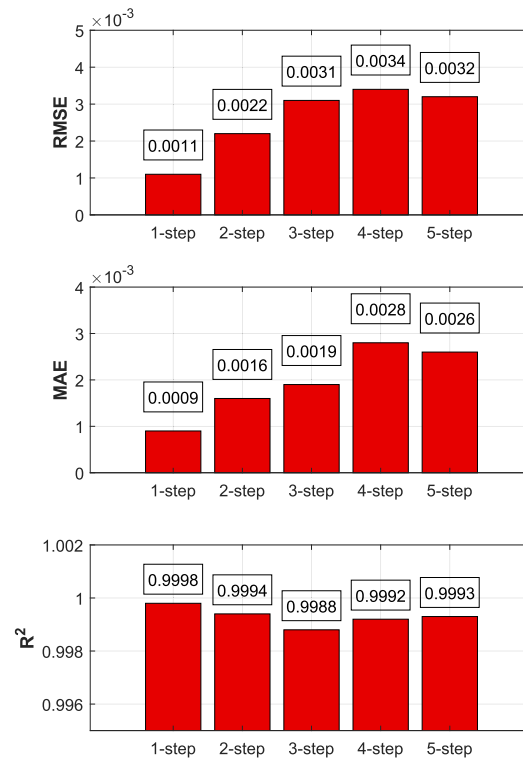


Fig. 14. Accuracy metrics for prediction horizons from 1 to 5 time steps.

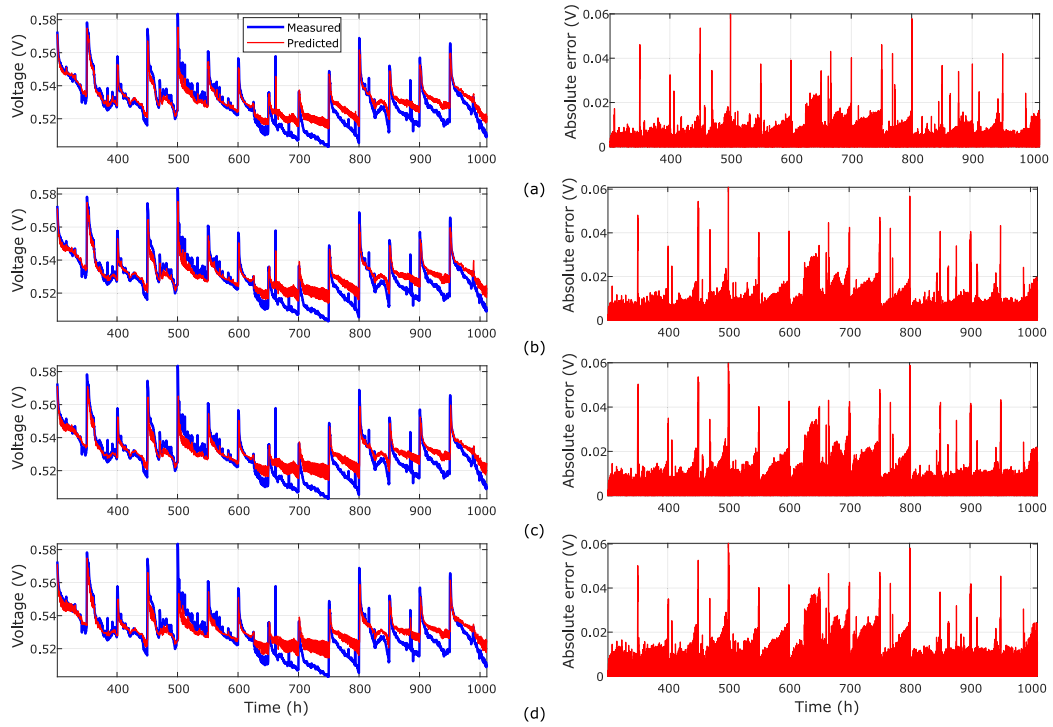


Fig. 15. Simulation results for Long-Term prognostic with increasing prediction horizons: (a) $PH = 70$, (b) $PH = 105$, (c) $PH = 140$, and (d) $PH = 175$.

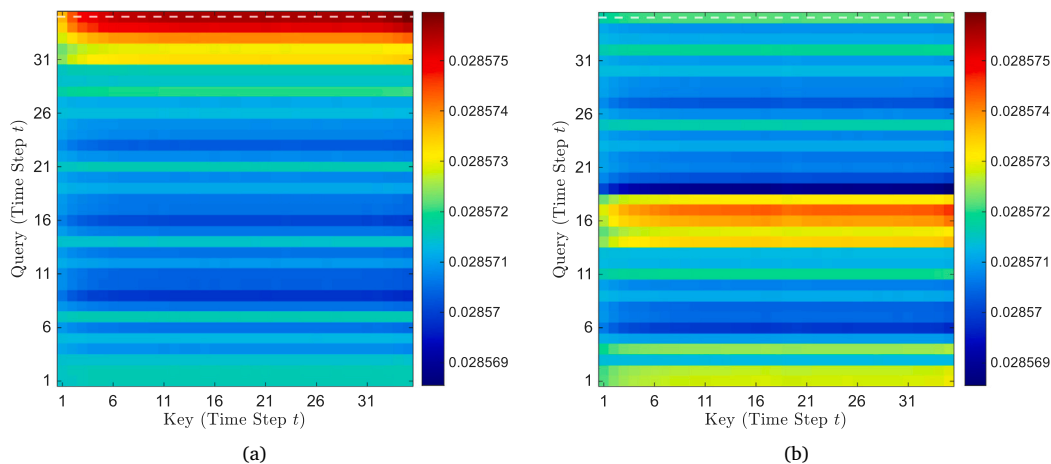


Fig. 16. Global map of the attention mechanism: (a) Low Current ($I = 1.76$ A); (b) Global attention map during High Current ($I = 35.54$ A), showing concentration on recent steps for reactive tracking.

CRedit authorship contribution statement

L.M. Perez: Writing – original draft, Visualization, Software, Methodology, Investigation, Conceptualization. **Mohsen Kandideyeni:** Writing – review & editing, Visualization, Supervision, Resources, Formal analysis, Conceptualization. **Alexandre Ravay:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation. **Javier Solano:** Writing – review & editing, Visualization, Supervision, Project administration. **Samir Jemei:** Writing – review & editing, Visualization, Supervision, Funding acquisition. **Loïc Boulon:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been supported by the EIPHI Graduate School (contract ANR-17-EURE-0002) and the Region Bourgogne Franche-Comté.

This work has also been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN-2025-05444.

References

- [1] X. Sun, D. Xi, J. Fu, H. Tian, W. Long, Boosted deep neural network model for forecasting the electrochemical impedance of a proton exchange membrane fuel cell under varying operating conditions, *Renew. Energy* 256 (2026) 124099, <http://dx.doi.org/10.1016/j.renene.2025.124099>.
- [2] C. Xu, A.N. Abdalla, Coordinated dispatch of electric, thermal, and hydrogen vectors in renewable-enriched microgrids using constrained harris hawks optimization under uncertainty, *Renew. Energy* 256 (2026) 124064, <http://dx.doi.org/10.1016/j.renene.2025.124064>.
- [3] Q. Zhang, D. Xie, Y. Zeng, Y. Liu, H. Yu, S. Liu, Optimizing wind-solar hydrogen production through collaborative strategy with ALK/PEM multi-electrolyzer arrays, *Renew. Energy* 232 (2024) 121116, <http://dx.doi.org/10.1016/j.renene.2024.121116>.
- [4] H. Lu, H. Huang, S. Zhang, S. Li, X. Guo, K. Xing, L. Tang, Hybrid time series-Exponential feature model for PEMFC degradation prediction, *Renew. Energy* 256 (2026) 124439, <http://dx.doi.org/10.1016/j.renene.2025.124439>.
- [5] International Energy Agency (IEA), *Global hydrogen review 2024*, 2024, Licence: CC BY 4.0.
- [6] International Energy Agency, *CO₂ Emissions from Fuel Combustion*, International Energy Agency, Paris, France, 2019, p. 165.
- [7] T. Abedin, J. Pasupuleti, J.K.S. Paw, Y.C. Tak, M. Mahmud, M.P. Abdullah, M. Nur-E-Alam, Proton exchange membrane fuel cells in electric vehicles: Innovations, challenges, and pathways to sustainability, *J. Power Sources* 640 (2025) 236769, <http://dx.doi.org/10.1016/j.jpowsour.2025.236769>.
- [8] P. Sharma, M. Cirrincione, A. Mohammadi, G. Cirrincione, R.R. Kumar, An overview of Artificial Intelligence-based techniques for PEMFC system diagnosis, *IEEE Access* (2024) <http://dx.doi.org/10.1109/ACCESS.2024.3493620>.
- [9] A. Pramuanjaroenikij, S. Kakaç, The fuel cell electric vehicles: The highlight review, *Int. J. Hydrog. Energy* 48 (2023) 9401–9425, <http://dx.doi.org/10.1016/j.ijhydene.2022.11.103>.
- [10] Z. Hua, Z. Zheng, E. Pahan, M.-C. Péra, F. Gao, A review on lifetime prediction of proton exchange membrane fuel cells system, *J. Power Sources* 529 (2022) 231256, <http://dx.doi.org/10.1016/j.jpowsour.2022.231256>.
- [11] V.B. Kumar, A. Collin, M.G. Sankar, K. Subramanian, Fuel cell technologies in the automotive sector: A focus on proton exchange membrane and Alkaline fuel cells, *Green Technol. Sustain.* 3 (2025) 100218, <http://dx.doi.org/10.1016/j.grets.2025.100218>.
- [12] M. Yue, S. Jemei, N. Zerhouni, R. Gouriveau, Proton exchange membrane fuel cell system prognostics and decision-making: Current status and perspectives, *Renew. Energy* 179 (2021) 2277–2294, <http://dx.doi.org/10.1016/j.renene.2021.08.045>.
- [13] L. Wang, X. Li, P. Guo, S. Guo, Z. Yang, P. Pei, Bibliometric analysis of prognostics and health management (PHM) in hydrogen fuel cell engines, *Int. J. Hydrog. Energy* 47 (2022) 34216–34243, <http://dx.doi.org/10.1016/j.ijhydene.2022.08.024>.
- [14] E. Wallnöfer-Ogris, F. Poimer, R. Köll, M.-G. Macherhammer, A. Trattner, Main degradation mechanisms of polymer electrolyte membrane fuel cell stacks – Mechanisms, influencing factors, consequences, and mitigation strategies, *Int. J. Hydrog. Energy* 50 (2024) 1159–1182, <http://dx.doi.org/10.1016/j.ijhydene.2023.06.215>.
- [15] J. Zuo, N.Y. Steiner, Z. Li, D. Hissel, Health management review for fuel cells: Focus on action phase, *Renew. Sustain. Energy Rev.* 201 (2024) <http://dx.doi.org/10.1016/j.rser.2024.114613>.
- [16] E.M. Koltsiva, V.A. Vasilenko, A.I. Shcherbakov, E.A. Fokina, V.A. Bogdanovskaya, Mathematical simulation of PEMFC platinum cathode degradation accounting catalyst's nanoparticles growth, *Chem. Eng. Trans.* 70 (2018) 1303–1308, <http://dx.doi.org/10.3303/CET1870218>.
- [17] Z. Zhang, H. He, Y. Wang, S. Quan, J. Chen, R. Han, A novel generalized prognostic method of proton exchange membrane fuel cell using multi-point estimation under various operating conditions, *Appl. Energy* 357 (2024) 122519, <http://dx.doi.org/10.1016/j.apenergy.2023.122519>, Cited by: 0.
- [18] J. Bartlechner, M. Vrljić, C. Hametner, S. Jakubek, State-of-health observer for PEM fuel cells—A novel approach for real-time online analysis, *Int. J. Hydrog. Energy* (2024) <http://dx.doi.org/10.1016/j.ijhydene.2024.03.061>.
- [19] L. Vichard, N.Y. Steiner, N. Zerhouni, D. Hissel, Hybrid fuel cell system degradation modeling methods: A comprehensive review, *J. Power Sources* 506 (2021) 230071, <http://dx.doi.org/10.1016/j.jpowsour.2021.230071>.
- [20] H. Liu, J. Chen, D. Hissel, H. Su, Short-term prognostics of PEM fuel cells: A comparative and improvement study, *IEEE Trans. Ind. Electron.* 66 (2019) 6077–6086, <http://dx.doi.org/10.1109/TIE.2018.2873105>.
- [21] K. Benagounne, M. Yue, S. Jemei, N. Zerhouni, A data-driven method for multi-step-ahead prediction and long-term prognostics of proton exchange membrane fuel cell, *Appl. Energy* 313 (2022) <http://dx.doi.org/10.1016/j.apenergy.2022.118835>.

- [22] T.J. Chang, S.J. Cheng, C.H. Hsu, J.M. Miao, S.F. Chen, Prognostics for remaining useful life estimation in proton exchange membrane fuel cell by dynamic recurrent neural networks, *Energy Rep.* 8 (2022) 9441–9452, <http://dx.doi.org/10.1016/j.egy.2022.07.032>.
- [23] Z. Tao, C. Zhang, J. Xiong, H. Hu, J. Ji, T. Peng, M.S. Nazir, Evolutionary gate recurrent unit coupling convolutional neural network and improved manta ray foraging optimization algorithm for performance degradation prediction of PEMFC, *Appl. Energy* 336 (2023) <http://dx.doi.org/10.1016/j.apenergy.2023.120821>.
- [24] Z. Liu, S. Xu, H. Zhao, Y. Wang, Durability estimation and short-term voltage degradation forecasting of vehicle PEMFC system: Development and evaluation of machine learning models, *Appl. Energy* 326 (2022) <http://dx.doi.org/10.1016/j.apenergy.2022.119975>.
- [25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: 31st Conference on Neural Information Processing Systems, Long Beach, 2017, pp. 1–15.
- [26] J. Lv, X. Shen, Y. Gao, J. Liu, G. Sun, The seasonal-trend disentangle based prognostic framework for PEM fuel cells, *Renew. Energy* 228 (2024) <http://dx.doi.org/10.1016/j.renene.2024.120648>.
- [27] J. Zuo, H. Lv, D. Zhou, Q. Xue, L. Jin, W. Zhou, D. Yang, C. Zhang, Deep learning based prognostic framework towards proton exchange membrane fuel cell for automotive application, *Appl. Energy* 281 (2021) 115937, <http://dx.doi.org/10.1016/j.apenergy.2020.115937>.
- [28] L. Hongwei, Q. Binxin, H. Zhicheng, L. Junnan, Y. Yue, L. Guolong, An interpretable data-driven method for degradation prediction of proton exchange membrane fuel cells based on temporal fusion transformer and covariates, *Int. J. Hydrog. Energy* 48 (2023) 25958–25971, <http://dx.doi.org/10.1016/j.ijhydene.2023.03.316>.
- [29] Y. Yu, Q. Yu, R.S. Luo, S. Chen, J. Yang, F. Yan, Deep learning with dual-stage attention mechanism for interpretable prediction of proton exchange membrane fuel cell performance degradation, *Int. J. Hydrog. Energy* 58 (2024) 902–911, <http://dx.doi.org/10.1016/j.ijhydene.2024.01.308>.
- [30] Y. Pan, H. Wang, N.P. Brandon, Gas diffusion layer degradation in proton exchange membrane fuel cells: Mechanisms, characterization techniques and modelling approaches, *J. Power Sources* 513 (2021) <http://dx.doi.org/10.1016/j.jpowsour.2021.230560>.
- [31] P. Ren, P. Pei, Y. Li, Z. Wu, D. Chen, S. Huang, Degradation mechanisms of proton exchange membrane fuel cell under typical automotive operating conditions, *Prog. Energy Combust. Sci.* 80 (2020) <http://dx.doi.org/10.1016/j.pecs.2020.100859>.
- [32] Z. Deng, Q. Chen, L. Zhang, K. Zhou, Y. Zong, H. Liu, J. Li, L. Ma, Degradation prediction of PEMFCs using stacked echo state network based on genetic algorithm optimization, *IEEE Trans. Transp. Electrification* 8 (2022) 1454–1466, <http://dx.doi.org/10.1109/TTE.2021.3111906>.
- [33] H.L. Nguyen, J. Han, H.N. Vu, S. Yu, Investigation of multiple degradation mechanisms of a proton exchange membrane fuel cell under dynamic operation, *Energies* 15 (2022) <http://dx.doi.org/10.3390/en15249574>.
- [34] M. Yue, Z. Li, R. Roche, S. Jemei, N. Zerhouni, Degradation identification and prognostics of proton exchange membrane fuel cell under dynamic load, *Control Eng. Pract.* 118 (2022) <http://dx.doi.org/10.1016/j.conengprac.2021.104959>.
- [35] H. Wu, W. Wei, Y. Li, W. Zhu, C. Xie, H.B. Gooi, Hybrid physics-based and data-driven prognostic for PEM fuel cells considering voltage recovery, *IEEE Trans. Energy Convers.* 39 (2024) 601–612, <http://dx.doi.org/10.1109/TEC.2023.3311460>.
- [36] H. Liu, J. Chen, D. Hissel, H. Su, Remaining useful life estimation for proton exchange membrane fuel cells using a hybrid method, *Appl. Energy* 237 (2019) 910–919, <http://dx.doi.org/10.1016/j.apenergy.2019.01.023>.
- [37] Y. Zhou, A. Ravey, M.C. Péra, Real-time cost-minimization power-allocating strategy via model predictive control for fuel cell hybrid electric vehicles, *Energy Convers. Manage.* 229 (2021) <http://dx.doi.org/10.1016/j.enconman.2020.113721>.
- [38] J. Zuo, H. Lv, D. Zhou, Q. Xue, L. Jin, W. Zhou, D. Yang, C. Zhang, Long-term dynamic durability test datasets for single proton exchange membrane fuel cell, *Data Brief* 35 (2021) 106775, <http://dx.doi.org/10.1016/j.dib.2021.106775>.
- [39] Y. Yang, Y. Yang, S. Zhou, H. Li, W. Zhu, Y. Liu, C. Xie, R. Zhang, Degradation prediction of proton exchange membrane fuel cell based on mixed gated units under multiple operating conditions, *Int. J. Hydrog. Energy* 67 (2024) 268–281, <http://dx.doi.org/10.1016/j.ijhydene.2024.04.186>.
- [40] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (8) (1997) 1735–1780, <http://dx.doi.org/10.1162/neco.1997.9.8.1735>.
- [41] S. Li, W. Luan, C. Wang, Y. Chen, Z. Zhuang, Degradation prediction of proton exchange membrane fuel cell based on Bi-LSTM-GRU and ESN fusion prognostic framework, *Int. J. Hydrog. Energy* 47 (2022) 33466–33478, <http://dx.doi.org/10.1016/j.ijhydene.2022.07.230>.