Sustainable Machine Intelligence Journal

Enhancing Prognostics of PEM Fuel Cells with a Dual-Attention LSTM Network for Remaining Useful Life Estimation: A Deep Learning Model

Ahmed Darwish [1,*]

[1] Department of Computer Science, Faculty of Computer and Informatics, Zagazig University, Zagazig, 44519, Egypt; adarwish@fci.zu.edu.eg

Event	Date
Received	09-02-2024
Revised	25-03-2024
Accepted	24-04-2024
Published	30-05-2024

Abstract: The Proton Exchange Membrane Fuel Cell (PEMFC) presents itself as an effective and viable technology to consider for transportation purposes. One of the most important things to consider in the field of electric vehicles is the crucial evaluation of the deterioration of the PEMFC stack. A data-driven deep learning method is proposed in this paper to enhance the precision of PEMFC Remaining Useful Life (RUL) prediction. This method combines a Long Short-Term Memory (LSTM), self-attention, and scaled dot-product attention mechanism. The LSTM enables the model to comprehend intricate temporal patterns and produce more abstract data representations. The correlations among various time points are found by using the self-attention mechanism. To highlight the most important features, the scaled dot-product attention method is used. In order to demonstrate the efficacy of the proposed model, comparisons with several Deep Learning models were made using the Dataset of the 2014 PHM Data Challenge. The findings from the experiments suggest that DA-LSTM proves to be a reliable choice for the RUL prediction of PEMFCs, as it demonstrated superior performance compared to all other models examined. The source code is available here: https://github.com/AhmedHossam10/remaining-useful-life-prediction-for-proton-exchange-membrane-fuel-cells

Keywords: Remaining useful life; Proton exchange membrane fuel cell; Long short-term memory network; Attention mechanism.

1. Introduction

These days, the automotive industry is confronted with two prominent global issues, namely global warming, and the energy crisis [1], [2]. Despite the utilization of advanced technologies such as turbocharging, direct injection, and Atkinson cycle engines to reduce pollutant emissions in gasoline-powered vehicles [3], [4], significant levels of CO2 emissions persist, contributing to the exacerbation of global warming. On the other hand, the full-scale adoption of electric vehicles faces challenges in the short term due to their limited range, slow charging times, and the lack of a comprehensive charging infrastructure [5]. Furthermore, it is worth noting that electric vehicles continue to generate CO2 emissions during their production phase when viewed from a life cycle standpoint [6]. Conversely, hydrogen fuel cell vehicles are emerging as a prominent avenue for the advancement of new energy transportation due to their environmental friendliness, rapid refueling capabilities, and extended driving range [7].

A hydrogen fuel cell vehicle's main component is a proton exchange membrane fuel cell (PEMFC). The chemical energy contained in a hydrogen tank is quickly converted into electrical energy via an electrochemical process. Its primary use

is in the realm of transportation due to its superior density of power and more dynamic properties than other alternative fuel cell varieties. These advantages stem from its key features, such as increased density of power, appropriate energy capacity range (0.001-100 kW), a narrow working temperature varying from 0 °C to 80 °C, lack of electrolyte loss, and lightweight design [8]. The two primary tasks in prognostics and health management involve forecasting degradation patterns and determining the remaining useful life (RUL). A key factor in increasing the efficiency of hydrogen fuel cell cars is the precise estimation of RUL for PEMFCs. Hence, it is imperative to develop a precise and efficient predictive model for RUL. The degradation pattern of PEMFCs may be assessed by using stack aging indices like as total electrochemical impedance spectroscopy (EIS), voltage, polarization curves, and power. Similarly, various parameters of degradation models derived from measurement data can serve as degradation indices. At present, the stack power and stack voltage are the most widely used measurement-based deterioration indicators for PEMFCs.

Lately, Scholars have recently devised a range of research methodologies aimed at forecasting proton exchange membrane fuel cell RULs. These approaches are typically categorized into three primary groups: model-based methods, data-driven methods, and hybrid methods [9] [10]. The model-based methodologies concentrate on constructing a theoretical method that elucidates how the fuel cell system operates. They posit that mathematical models or formulae can accurately represent the system's degradation. Model-based methodologies strive to provide accurate RUL estimations based on fundamental physics and chemistry that are relevant to fuel cells by clarifying the degradation mechanisms and their impact on the system [11]. The degradation phenomenon of PEMFCs is simulated through model-based methods, utilizing mechanism models that do not rely heavily on data. To predict the prognosis of PEMFCs, various models have been used, including the empirical deterioration method [12] [13], mechanism deterioration method [14], semi-empirical deterioration method [7] [8], and semi-mechanism deterioration method [17]. Chen et al. [18] introduced a technique for real-time forecasting of the RUL of PEMFCs utilizing data from laboratory testing and vehicle operations, serving as a foundation for PEMFC lifespan design. An expanded Kalman filter technique and a voltage aging model were used in Chen et al.'s [19] approach for predicting aging. By using deterioration methods for equivalent resistance and electrochemical surface area, Ouyang et al. [20] developed a voltage prediction method to forecast the course of PEMFC deterioration. Lechartier et al. [21] developed a predictive framework for PEMFCs, comprising a distinct static component and a dynamic component, and verified its accuracy using empirical data obtained from extended testing periods. Moreover, the process of developing precise physical models through trial and error to match experimental results can be a lengthy process lasting for multiple years. Data-driven methods involve constructing predictive models using historical data obtained from sensors on machinery and the associated measurements. The application of data-driven methodologies is effective in achieving generalization without requiring specialized expertise. Through the use of data-driven techniques, the relationship between sensor data and system degradation can be uncovered [22], demonstrating a robust ability to generalize and reduce reliance on empirical knowledge [23]. Data-driven approaches for RUL have the potential to lessen the reliance on conventional engineering knowledge and generate accurate predictions by examining trends in temporal data. This approach is advantageous in facilitating the ongoing, live tracking of equipment, thus decreasing the likelihood of crucial malfunctions while also improving forthcoming production procedures [24]. Data-driven methodologies in predicting RUL for PEMFCs entail utilizing historical and real-time data for detecting significant patterns and trends. Subsequently, a black-box model was provided with this information as input, which was then analyzed and converted into pertinent measurements. Creating the connection between the intended output (RUL prediction) and the input sensor monitoring data (For instance, environmental or operational data) is largely dependent on this black-box model. It discerns and encapsulates relationships and latent information from the data for facilitating precise forecasting.

This study evaluates the RUL of complex systems by utilizing data-driven methodologies. These techniques encompass Deep Learning (DL) and Machine Learning (ML) approaches. ML methodologies have become a powerful tool in various aspects of our daily lives, enabling computers to learn from data autonomously and improve their effectiveness continuously without explicit programming. The capacity of ML to extract insights from data and perform tasks automatically is transforming our way of life, work practices, and technological engagements. With the advancement of this field, we anticipate even greater influences on our global community. ML methodologies can leverage large volumes of sensor data, operational parameters, and historical maintenance records. Through this methodical data-driven approach, machine learning algorithms can grasp intricate relationships among diverse variables that influence the state and deterioration of equipment. Complex feature extraction is conducted across different domains like frequency, time, and time-frequency [25], and the selection of features is crucial for ensuring the optimal performance of traditional machine learning methods. hybrid deterioration tracking methods and Support Vector Machines (SVMs) were combined by Yan et al. The SVM was employed for the classification of degradation across five stages, while the hybrid deterioration tracking method was utilized to determine the most suitable RUL values for each stage of degradation [26]. Wavelet packet decomposition was used by Mejia et al. [27] to acquire coefficients from the primary sensor data. This entailed evaluating the current operational condition of the equipment using a Gaussian Hidden Markov Model (HMM). Additionally, the HMM model was utilized to predict the machinery's RUL along with determining the associated level of confidence. An intermediate domain was introduced by Shen et al. [28] to address the issue of source degradation indices being inadequate for SVM applications. In their approach they choose the best feature degradation indices, they employ a combination of PCA and the joint evaluation index. When it came to predicting RUL with little data, Nan et al.'s combination of gray predictive modeling (GM) and Relevance Vector Machine (RVM) with dynamic window widths demonstrated superior performance better than both a Convolutional Neural Network (CNN) and a Particle Filter (PF). The GM can anticipate the trend of deterioration with the use of relevance vectors generated by the RVM. In this methodology, dynamic window sizes are implemented to cater to various life cycle stages, wherein windows with longer widths exhibit more efficient performance in the early stages while windows with smaller widths are more effective in the later stages [29]. Zhang et al. utilized the RVM in conjunction with the differential evolution (DE) algorithm for predicting the batteries' RUL by utilizing denoised data generated through the wavelet denoising algorithm [30]. Tang et al. developed a collection of RVMs in which the primary model is adapted to specific characteristics, while each additional RVM is adapted to an individual feature. In addition to employing a comprehensive approach to feature generation through various techniques, they also introduced a method for feature selection to prioritize the features based on their significance over time [31].

Deep learning, a notable subset within the domain of machine learning, has led to significant changes in various facets of our society. It combines artificial neural networks with multiple layers to process data, mimicking the cognitive processes of the human brain. The advancements in deep learning methods have been impressive due to their versatility. Unlike ML, deep learning models do not require feature engineering as they can independently extract feature representations. Deep learning stands out for its complex network architecture, facilitating the identification of deteriorating features from past sensor data of monitored equipment [32]. As a result, deep learning demonstrates superior effectiveness in managing unstructured and high-dimensional data types like time series data and images. Notably, deep learning methods surpass other machine learning techniques, particularly in the domain of predictive maintenance. Wang et al. [33] presented the spatiotemporal non-negative projected convolutional network (SNPCN) methodology for detecting deterioration patterns within adjacent matrices by employing a three-dimensional convolutional neural network (3DCNN). The authors later used the PRONOSTIA platform to validate the efficacy of

this approach. Wang et al. introduced a Temporal Convolutional Network (TCN) that integrates two distinct convolution techniques to enhance the capacity for capturing local dependencies and enhancing accuracy: causal convolutions, which enable the examination of preceding sequential steps, and dilated convolutions, which facilitate the observation of a broader segment of the entire sequence [34]. Similarly, Yang et al. [35] devised a dual-CNN model for the prediction of RUL. The primary CNN is devoted to identifying the initial point of failure, while the secondary CNN is employed to predict the RUL value. Using a special version of the recurrent neural network (RNN) called grid-long short-term memory (G-LSTM), Ma et al. [36] presented a novel technique for forecasting fuel cell deterioration. The accuracy of the model was verified by testing using three distinct kinds of PEMFCs. A study [37] suggested an Echo State Network-based (ESN-based) method for prognosticating fuel cells under varying loads, with a notably low level of uncertainty in the estimation method. Wang et al. [38] suggested an integration of an attention mechanism within a bidirectional LSTM model to conduct RUL regression specifically for lithium batteries. In a similar vein, Chen et al. [39] presented a methodology that encompasses a two-step process first, for feature extraction they used Kernel Principal Component Analysis (KPCA), and second, for RUL estimation they used an application of a Gated Recurrent Unit (GRU) based architecture.

The degradation system model is first built using a model-based strategy, which is based on a comprehensive comprehension of the behavior and degradation mechanisms of the system. The chemical and physical mechanisms that contribute to deterioration are included in this model. Next, utilizing the generated model and accessible data, a data-driven strategy is used to estimate the RUL. Hybrid methods use the best features of both approaches to provide more precise and dependable RUL forecasts. Using the benefits of both the data-driven strategy which uses actual data and statistical techniques for improving the accuracy of RUL prediction and the model-based approach which integrates fundamental physics and chemical concepts into the fuel cell system the hybrid approach maximizes its potential. Through the integration of these methods, a deeper understanding of the performance of the system and deterioration trends may be attained, resulting in an enhanced estimation of RUL [21]. Cheng et al. [40] suggested a new method for predicting the PEMFCs' performance, utilizing a combination of the regularized particle filter (RPF) and the least square support vector machine (LSSVM). By applying this approach to the FCLAB Research Consortium's PEMFC dataset, the effectiveness of this methodology was confirmed. Meanwhile, Liu et al. [41] employed a data-driven method using an automated ML technique that utilized the adaptive neuro-fuzzy inference system (ANFIS). To predict the remaining operational life, they also used an adaptive centerless Kalman filter (ACKF) technique in combination with a semi-empirical deterioration method. The Least Squares SVM (LSSVM) is applied in various areas within the field of PHM. A technique called a dual filter approach was presented by Li et al. [42] that combines LSSVM with Unscented Particle Filters (UPF). The real-time updating of degradation states and LSSVM parameters is made possible by this fusion method, which combines model-based and data-driven elements. As a UPF's measurement equation, the LSSVM model which is first trained offline generates predictive values for future observations. Additionally, the LSSVM training set receives feedback from the expected output and LSSVM model parameter outcomes to enable dynamic parameter modifications and live training. Mechanism models are usually necessitated for both hybrid and model-based techniques, which leads to intricate modeling and sophisticated computing. Consequently, multiple researchers are focusing more on data-driven techniques, which encounter the obstacle of enhancing the model's precision accuracy.

We introduce a novel data-driven deep learning method in this study aimed at predicting the RUL of PEMFCs. A novel technique referred to as dual attention combined with a long short-term memory (DA-LSTM) network is suggested in order to forecast RUL by taking sensor data obtaining temporal dependencies and extracting spatial

characteristics. An attention mechanism has been integrated into the LSTM architecture in order to allocate weight values to the extracted features, thereby emphasizing crucial information and enhancing the model's RUL forecast. The DA-LSTM model's efficiency is assessed using the well-known IEEE PHM 2014 Challenge Dataset [43] for predicting the RUL of PEMFCs. Our experimental findings demonstrate that the proposed approach diminishes uncertainty in multi-step prediction tasks and surpasses other existing models in terms of accuracy. The study's noteworthy contributions can be summarized as follows:

1) A novel data-driven deep learning architecture designed to understand temporal relationships in time series data. This is achieved through the integration of dual attention mechanisms with long short-term memory (LSTM), which efficiently captures deterioration patterns in the PEMFCs.
2) The model's validity was verified through the utilization of the dataset of the IEEE 2014 PHM data challenge [43], demonstrating its superior performance in predicting RUL compared to other models.

The following sections of this document are organized in a subsequent manner. Section 2 includes the description of Methods and materials. Following this, Section 3 provides detailed information on the suggested methodology. This is then succeeded by the elucidation of Experimental settings in Section 4. Section 5 delves into the Results and discussion. Lastly, Section 6 summarizes the conclusions from this study.

2. Methods and materials

2.1 Long Short-Term Memory (LSTM)

The network can operate similarly to a series of hidden layers interacting with each other because the Recurrent Neural Network (RNN) adds a feedback loop to the hidden layer. RNNs are well-suited for the analysis of temporal data because of this particular feature that allows them to remember past events. RNNs encounter two potential limitations when utilizing gradient-based backpropagation to modify the network weights. Neural networks may encounter issues with the disappearing or expanding gradient as the number of hidden layers increases, especially when generating predictions of long-term [44]. The necessity to examine extended periods of time intervals in forecasting necessitated the exploration of alternative approaches beyond RNNs to ensure precise outcomes. Various iterations of RNNs have been developed, such as the special version of RNN called Long Short-Term Memory.

A novel RNN architectural modification is presented by the LSTM, specifically crafted to address the issue of disappearing gradients in conventional RNNs, with a focus on resolving the challenges posed by long-term dependencies in predictive tasks. This innovative structure includes a sophisticated configuration of memory cells that stands out for its ability to effectively retain and utilize information over extended sequences, making it particularly suitable for tasks that involve predicting long-term dependencies, such as RUL prediction in PEMFCs. Figure 1 shows the construction of the LSTM cell. Within the LSTM model, the forget gate f_t, input gate i_t, and output gate o_t constitute the three essential components that govern information flow and regulate interactions within the network. The information from the previous cell state that should be discarded is determined by the forget gate f_t. Taking into account the prior hidden state and the input at hand, the forget gate f_t produces an output value within the range of 0 to 1, representing complete forgetfulness to full retention. The mathematical expression for f_t is detailed in Equation 1.

$$f_t = \sigma\left(W_f\, x_t + \, U_f\, h_{t-1} + \, b_f\right) \quad (1)$$

Where the sigmoid function is shown by the symbol σ, t represents the time step, the output hidden state from the previous time sample is indicated by h_{t-1}, the input feature at time t is indicated by x_t, the parameters W_f, U_f, b_f are optimized during the training process. The input gate i_t is tasked with selecting new data to retain in the cell state. Through an analysis of the current input and preceding hidden state, it produces an output between 0 and 1, for incorporation into the cell state, along with a fresh candidate value. Equation 2 defines the mathematical form of i_t.

$$i_t = \sigma(W_i\, x_t + U_i\, h_{t-1} + b_i) \quad (2)$$

Where the parameters W_i, U_i, b_i are optimized at the training process. With reference to the previous hidden state and current input, what data is conveyed as the current LSTM cell's hidden state is governed by the output gate o_t. Its output varies from 0 to 1. The mathematical formulation of o_t is presented in Equation 3.

$$o_t = \sigma(W_o\, x_t + U_o\, h_{t-1} + b_o) \quad (3)$$

Where the parameters W_o, U_o, b_o are optimized during the training process. The candidate value c'_t represents novel data that has the potential to be integrated into the cell state at the current time step (t), generated by taking into account the input gate's consideration of the prior hidden state and the present input. Equation 4 defines the mathematical form of c'_t.

$$c'_t = tanh(W_a\, x_t + U_a\, h_{t-1} + b_a) \quad (4)$$

Where the parameters W_a, U_a, b_a are optimized at the training process. Subsequently, the c_t value representing the unit state at time t is calculated using Equation 5. Finally, the h_t value representing the hidden state at time t is determined utilizing mathematical Equation 6.

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t \quad (5)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (6)$$

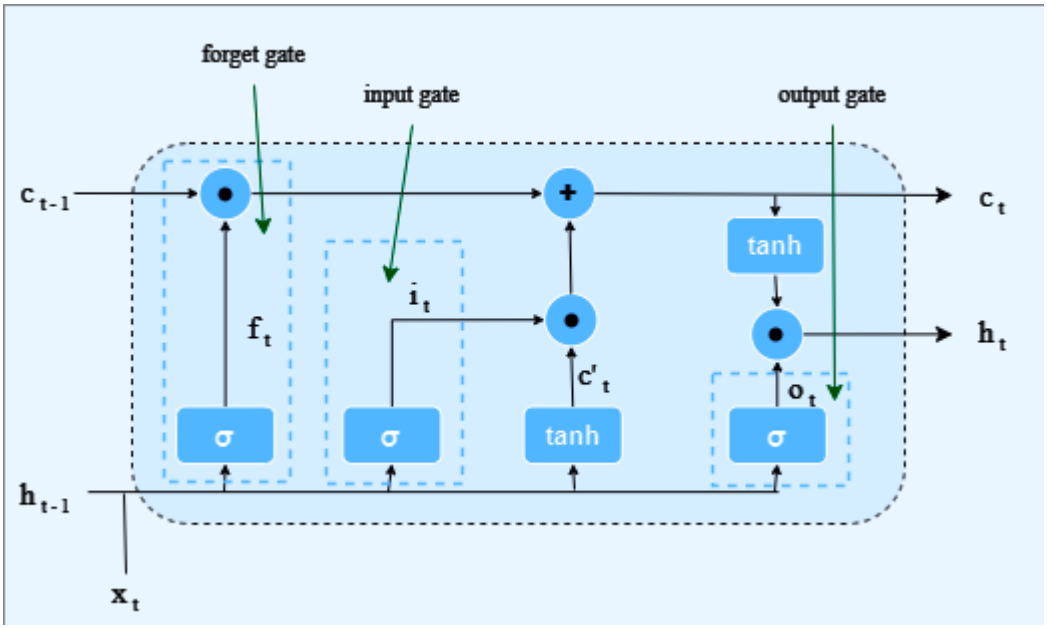

Figure 1. LSTM cell architecture.

2.2 Scaled dot-product attention

In the domain of time series analysis, recent data plays a pivotal role in accurately forecasting future values. Scaled dot-product attention is a method that helps highlight these recent portions of the LSTM output by giving more importance to their respective key vectors during the scoring phase. This technique is a key element used in different attention mechanisms. By incorporating it after the LSTM stage, The LSTM output sequence can include distinct segments that the model can concentrate on that are considered most relevant for prediction purposes. The transformation of the LSTM output involves mapping it into three separate vector spaces - Query, Key, and Value - using linear conversions. The query vector (Q) represents the center of attention for the model at each time step. The key vector (K) captures the available information at each time step in the sequence, while the value vector (V) holds the actual data content for each time step. A score is calculated by multiplying the query vector (Q) with each key vector (K) across the sequence, as shown in Equation 7.

$$score = \frac{Q \ . \ K^T}{\sqrt{d_k}} \quad (7)$$

Where d_k donates the dimensionality of the key vectors, and $\sqrt{d_k}$ represents a scaling factor to stabilize the gradients. Following this, the scores undergo a softmax function application to transform them into a probability distribution referred to as the attention weights (A) in Equation 8.

$$A = softmax(score) \quad (8)$$

According to the current prediction (which is based on the query vector), these weights show the relative relevance of each time step in the sequence. Finally, the attention weights (A) are used to assign importance to the corresponding value vectors (V) following Equation 9. This procedure produces a context vector that summarizes the most relevant information from the entire sequence based on the current focus (query).

$$context_{vector} = A \ . \ V \quad (9)$$

2.3 Self-attention mechanism

Traditional time series models frequently encounter challenges in capturing intricate relationships within the data, particularly when dealing with long-term dependencies. The use of self-attention helps in identifying dependencies across different time points, facilitating effective learning and representation of temporal relationships. Unlike traditional sequential data processing methods, self-attention allows each time point to consider all others in the sequence, enabling the model to select the most relevant past values for future predictions. Time series data often display prolonged dependencies, in which the measurement at a particular time instance is impacted by distant past data points. Self-attention mechanisms, specifically those integrating dilated convolutions or multi-head attention, are proficient in capturing these extended dependencies, thereby empowering the model to generate forecasts through a more comprehensive contextual comprehension. Through the utilization of the attention mechanism, information from individual time points is consolidated into the hidden state through the allocation of weights to the crucial data in the sequence. This mechanism functions by utilizing three vectors that are obtained by converting the input X into separate feature spaces: Keys (K), Queries (Q), and Values (V). The query vector (Q) symbolizes the central point of attention of the model at each time interval as determined through equation 10, while the key vector (K) contains the existing information at each time instance as computed by equation 11, and the value vector (V) retains the factual data content for each time step as calculated by equation 12.

$$Q = X \times W_q \quad (10)$$
$$K = X \times W_k \quad (11)$$
$$V = X \times W_v \quad (12)$$

where W_q, W_k, and W_v are trainable parameters, and X indicates the input tensor. The calculation involves evaluating the dot product between the query Q and each K in the sequence, then dividing by $\sqrt{d_k}$ after the dot product operation, with the dimension of K being specified. This process is illustrated in Equation 13.

$$attention_{scores} = \frac{Q \ . \ K^T}{\sqrt{d_k}} \quad (13)$$

Subsequently, the computed scores undergo a softmax transformation, resulting in a probability distribution known as attention weights (A) as described in Equation 14.

$$A = softmax(attention_{scores}) \quad (14)$$

These weights signify the importance of individual time steps in the sequence concerning the current prediction made by the query vector. Following this, the attention weights (A) are utilized to assign importance to the corresponding value vectors (V) based on Equation 15.

$$context_{vector} = A \ . \ V \tag{15}$$

This method leads to the formation of a context vector containing the most pertinent details from the entire sequence based on the current focus indicated by the query.

3. The proposed DL model

The RUL for PEMFCs is regarded as a supervised regression task, necessitating the utilization of data from the dataset of the Challenge of the IEEE 2014 PHM Data for training and assessing deep learning models. This study suggests a new data-driven deep learning method named DA-LSTM, which combines LSTM, self-attention, and a scaled dot-product attention mechanism for predicting the RUL of PEMFCs, as illustrated in Figure 2. To extract historical patterns of the supplied information in the temporal domain, the LSTM layer is employed. The result from the initial LSTM layer is then directed to the self-attention mechanism, responsible for recognizing dependencies among various time points. Subsequently, the LSTM layer transmits information through memory cells and gates. The output from the second LSTM layer serves as the input for the subsequent layer, aiding in capturing hierarchical representations of the input data. The neuron weights of the hidden state layer are subsequently ascertained by using the scaled dot-product attention method. This layer is crucial in computing the output via the LSTM, assigning weight coefficients, and reconstructing data to identify the essential features. Following is the last LSTM layer. The incorporation of multiple LSTM layers enables the model to grasp intricate temporal relationships and create more abstract data representations, incorporating dropout layers to prevent overfitting. The processed data from the hidden layer is combined and for the final RUL prediction, this processed data is directed to the fully connected (FC) layer. Lastly, Algorithm 1 outlines the pseudocode of the proposed model.

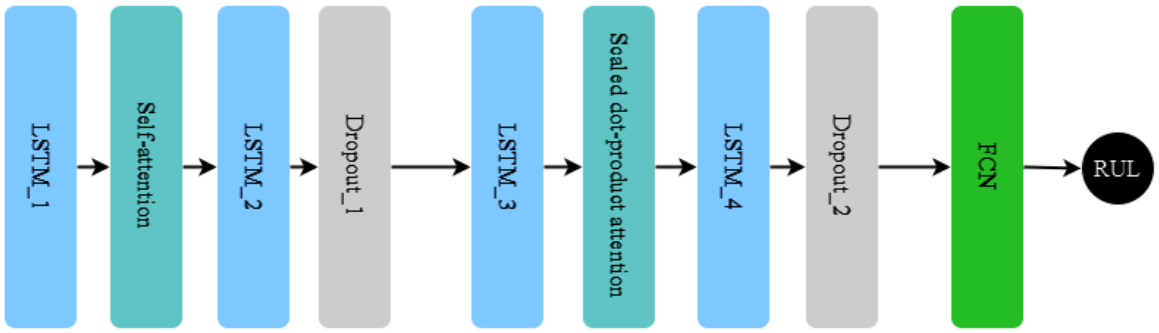

Figure 2. The suggested DA-LSTM flowchart.

Outlined in Algorithm 1, the suggested framework follows a series of steps to manage the input data. Initially, the input data undergoes preprocessing before being directed to the input layer. Subsequently, this information is fed into an LSTM with 256 neurons and a Tanh activation function to acquire temporal insights from the input data. The LSTM's resulting output is then passed to the self-attention mechanism to identify interdependencies among different time points. It is then sent to another LSTM layer with 256 neurons and a Tanh activation function, incorporating a dropout layer to prevent overfitting. Following this, an additional LSTM using 128 units and as an activation function using

Tanh is employed to allow the model to understand complex temporal patterns and generate more abstract data representations. The output of this layer is directed to the scaled-dot product attention mechanism dedicated to the key elements. Subsequently, it goes through another LSTM layer with 128 neurons and a Tanh activation function, along with the integration of a dropout layer. The resulting output is then channeled into a fully connected network that contains a dense layer using 1 unit serves as an output layer for predicting the remaining useful life of PEMFCs.

Algorithm 1 Pseudo-code of DA-LSTM

Input: Input data (D), batch size (Bs), maximum epoch (T), and learning rate (lr)
Output: *loss* (MSE), $RMSE$

Conducting the **preprocessing step**
/* Create the proposed DA-LSTM model */
Input: Construct an **input layer** to receive the input data
/* temporal learning based on the LSTM */
x: Create an **LSTM** layer with 256 units and a ***Tanh*** activation function to take the data from the **input layer.**
x: Add **Self-attention mechanism** to **x.**
x: Add **LSTM** layer with 256 units and ***Tanh*** activation function to **x.**
x: Add a **Dropout** layer with 0.4 value as a dropout rate to **x.**
x: Add an **LSTM** layer with 128 units and ***Tanh*** activation function to ***x.***
x: Add **Scled dot-product attention mechanism** to **x.**
x: Add an **LSTM** layer with 128 units and ***Tanh*** activation function to ***x.***
x: Add a **Dropout** layer with 0.4 value as a dropout rate to **x.**
/* Prediction Block */
x: Add a **dense layer** with **1** node to ***x.***
/* Optimization process */
$N = Size(D)/Bs$ /* Estimate the number of batches */
$t = 0$, Current epoch
while $t < T$
    $i = 0$, the current batch size.
    while $i < N$
        Compute the **Score function** using the ith **batch.**
        Optimize the **MSE** function by updating the weights according to Adam.
        $i = i + 1$
    end while.
    $t = t + 1$
end while

4. Experimental settings

4.1 2014 PHM Data Challenge Dataset

The dataset of the challenge of the IEEE PHM 2014 Data provides the data used in this study to analyze PEMFC [43]. As depicted in Figure 3, the FCLAB conducted the experimental examination. The fuel units appear in this research study in a stacked configuration, just like in many other research investigations. The system's power output and total voltage are enhanced by this stacking configuration. numerous fuel units linked in sequence can make up the stack, depending on the specific application's voltage and power requirements [45]. A stack of PEMFC with four fuel units positioned one above the other as shown in Figure 4. It visually demonstrates the stacking arrangement, highlighting the cells' series link. Furthermore, A basic design of the operational system is featured in the illustration, offering a summary of the way the fuel units fit within the complete structure. Five PEMFCs, especially of the Ulmer BZ 100 type, make up the stack of PEMFCs used in the experiment. The highest power output of the entire stack is 600 W. Table 1 outlines more details on the experimental operating settings and the features of the stack of PEMFC.

Table 1. Experimental operating conditions and features of PEFMC stacks.

Parameters	Value
Number of units	5
the active area of each cell	100 cm^2
Gas temperature	20 °C : 80 °C
Airflow	0: 100 1/min
Gas humidification	0 : 100 % RH
Cell standard voltage	0.6 V
H2 flow	0: 30 1/min
Gas pressure	0: 2 bars
Current	0: 300 A
Cooling flow	0: 10 1/min
Cooling temperature	20: 80 °C

Sensors installed at the PEMFC stack's intake and output measured a variety of degradation factors. Table 2 provides specifics for these criteria, offering valuable insights into the stack's condition and efficiency. In the data analysis process, the FC1 dataset was utilized for training and was fully accessible. Conversely, the FC2 experimental data, used for testing, was only available for the initial 550 hours, using the rest of the 551–1155 hours of data applied to RUL estimation. Figure 5 shows how the training and testing sets were distributed, outlining the study's data split.

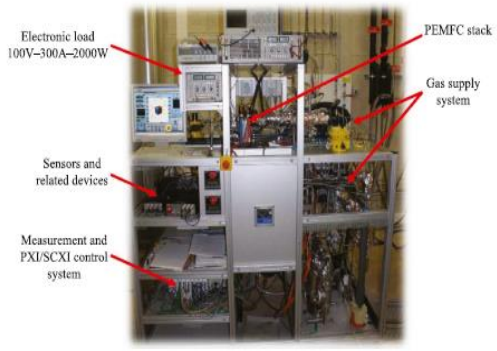

Figure 3. FCLAB degradation test platform.

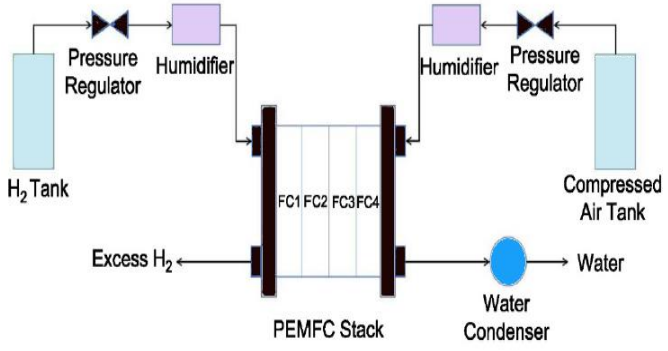

Figure 4. Schematic of PEMFCs system.

Table 2. IEEE 2014 PHM data challenge aging parameters.

parameter	Physical Meaning
Time	Aging time (h)
U_1 to U_5	Single-cell voltages (V)
U_{tot}	Stack voltage (V)
I	Current (A)
J	Current Density (A/cm^2)
TinH2	Inlet temperatures of H2 (°C)
ToutH2	outlet temperatures of H2 (°C)
TinAIR	Inlet temperatures of air (°C)
ToutAIR	outlet temperatures of air (°C)
TinWAT	Inlet temperatures of cooling water (°C)
ToutWAT	outlet temperatures of cooling water (°C)
PinH2	Inlet pressure of H_2(mbara)
PoutH2	outlet pressure of H_2(mbara)
PinAir	Inlet pressure of air(mbara)
PoutAir	outlet pressure of air(mbara)
DinH2	The inlet flow rate of H_2(l/mn)

DoutH2	outlet flow rate of H_2(l/mn)
DinAIR	Inlet flow rate of air (l/mn)
DoutAIR	outlet flow rate of air (l/mn)
DWAT	Flow rate of cooling water (l/mn)
HrAIRFC	Estimated air inlet hygrometry (%)

4.2 Data preprocessing

There are 127,370 samples in the dataset of FC2 and 143,862 samples in the dataset of FC1. To maintain the deteriorating features of the PEMFC and minimize the amount of computation required, every parameter as well as samples across the two sets was collected for one hour. As a consequence, the FC2 and FC1 sets produced, respectively, 1021 and 1155 samples. The FC1's stack voltage displays an erratic and turbulent trend, whereas the FC2's stack voltages exhibit strong peaks because of the changes in load circumstances. By adding a window size equal to 21 for the Savitzky-Golay filter (SGF) and applying it to the two stack voltage sets, we decreased the noise and softened the generated sets. The influence of SGF is illustrated in Figure 6 for FC1 and Figure 7 for FC2.

In order to normalize the characteristic values, the min-max scaling approach is used. In machine learning, this scaling technique is important, particularly for characteristics with various magnitudes or measures. It standardizes the range while adjusting the data to a constant range, usually between 0 and 1 [46]. This preserves the original distribution form. Mathematically, min-max scaling is defined by equation 16.

$$x'_{i,j} = \frac{x_{i,j} - x_{j\,min}}{x_{j\,max} - x_{j\,min}} \quad (16)$$

Where $x_{i,j}$ donates the value of ith sample, and jth feature, $x_{j\,min}$, $x_{j\,max}$ donates the minimum, and maximum values in jth feature, respectively.

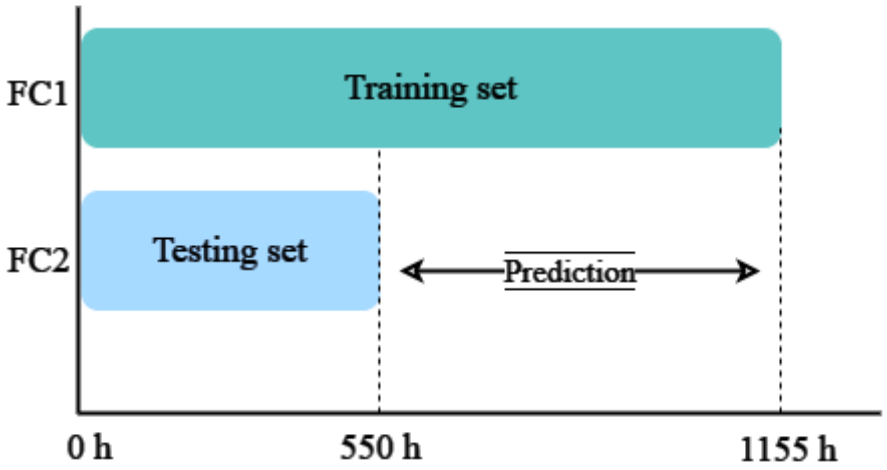

Figure 5. An instance of a dataset for testing and learning.

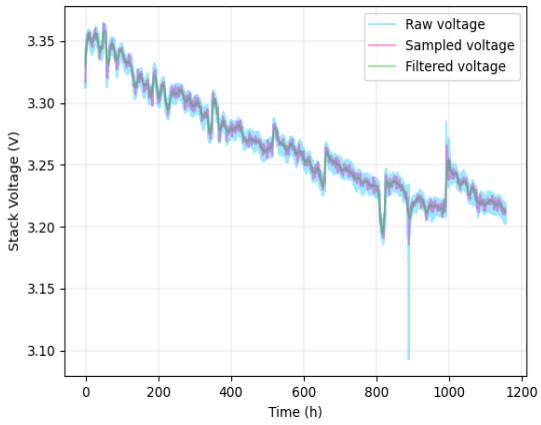

Figure 6. The influence of SGF on FC1.

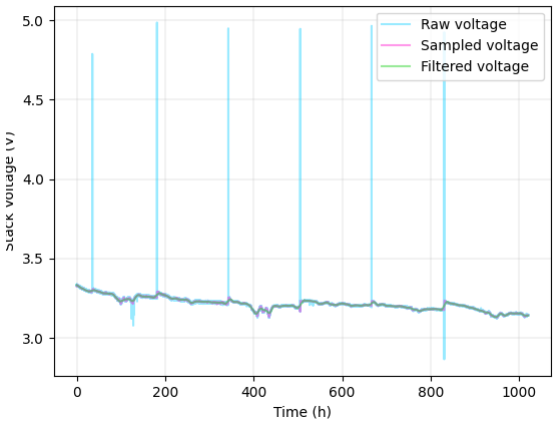

Figure 7. The influence of SGF on FC2.

4.3 Evaluation metrics

In this paper, the Adam optimization [47] technique and mean square error (MSE) loss, which can be determined using equation 17, are employed to enhance the network parameters.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - y'_i\right)^2 \quad (17)$$

Where N represents an overall sample size, also both the real and predicted labels of the ith sample are denoted by y_i and y'_i, respectively. The evaluation metric known as root mean square error (RMSE) has been used to appraise the proposed model by contrasting RMSE values between the real and forecasted labels of each data point in the dataset. The RMSE value is calculated through a mathematical formula specified in equation 18.

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

The reduction of both MSE and RMSE is essential for enhancing the accuracy in predicting the RUL of PEMFCs

4.3 Hyper parameter tuning

The DA-LSTM model proposed in this study incorporates several hyper-parameters, such as the number of units within each LSTM layer, the dropout rate, and the learning rate. It is essential to precisely define these parameters to enhance the model's effectiveness and decrease the RMSE. Therefore, Table 3 depicted that several experiments were conducted in this paper to explore different configurations for each parameter in order to identify the most optimal values that lead to a substantial boost in the performance of the model. The influence of these hyperparameters on the efficacy of the DA-LSTM approach is illustrated in Figures 8 to 14. For instance, the model's performance is significantly influenced by the quantity of units in each LSTM layer. Hence, multiple experiments were conducted to ascertain the best quantity of units for each LSTM layer, starting from 64, 128, to 256. The impact of the quantity of units for each LSTM layer can be observed in Figures 8, 9, 10, and 11. The dropout rate serves as a critical hyperparameter during the training of neural networks, playing a key role in mitigating overfitting. Overfitting is the phenomenon wherein a model not only captures the inherent patterns present in the training data but also incorporates noise and intricacies that do not translate effectively to new, unseen data. Many experiments are conducted to choose the best dropout rate. The impact of different dropout rates is visualized in Figure 12. The learning rate serves as a pivotal hyperparameter in the training of deep learning models, as it governs the magnitude of adjustments implemented in the optimization process for

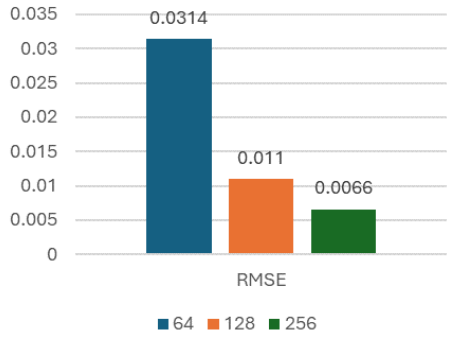

Figure 7. The impact of LSTM_1 unit number through experimentation.

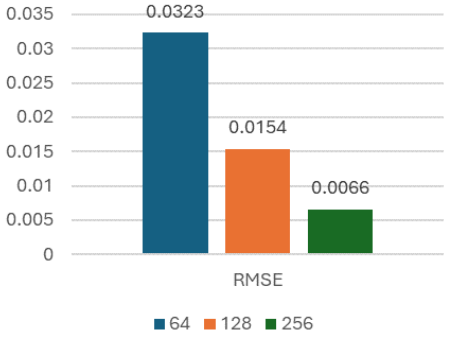

Figure 8. The impact of LSTM_2 unit number through experimentation.

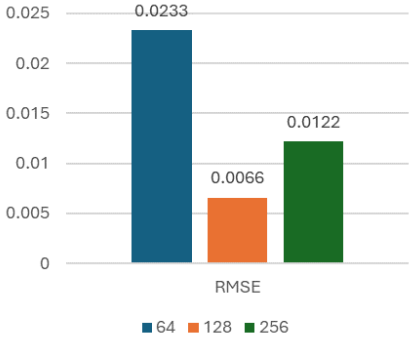

Figure 9. The impact of LSTM_3 unit number through experimentation.

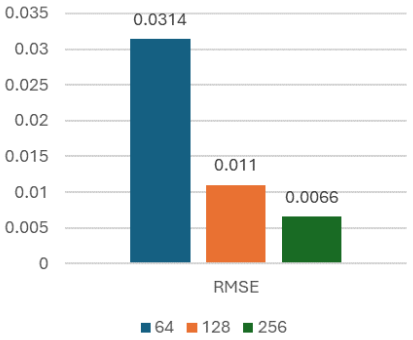

Figure 10. The impact of LSTM_4 unit number through experimentation.

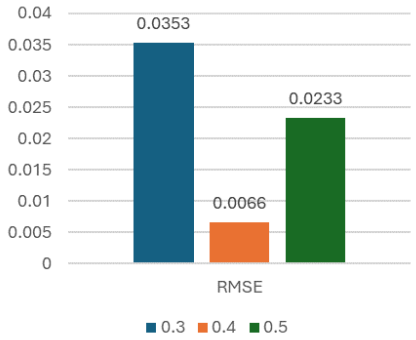

Figure 11. The influence of dropout rate through experiments.

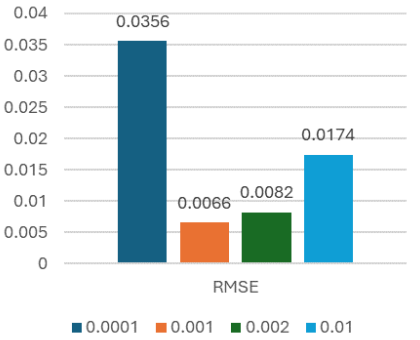

Figure 12. The influence of learning rate through experiments.

guiding the method's parameters toward minimizing the loss function. Various experiments were carried out to identify the most appropriate learning rate from the options of 0.0001, 0.001, 0.002, and 0.01, with results suggesting that the optimal learning rate is 0.001. The influence of different learning rate values is illustrated in Figure 13.

5.2 FC1 and FC2 prediction results

For the model that was trained on FC1, specific time intervals were selected for each stage, with 577 hours allocated for training, 115 hours for validating, and 462 hours for testing. Conversely, with FC2, the model trained has 510 hours for

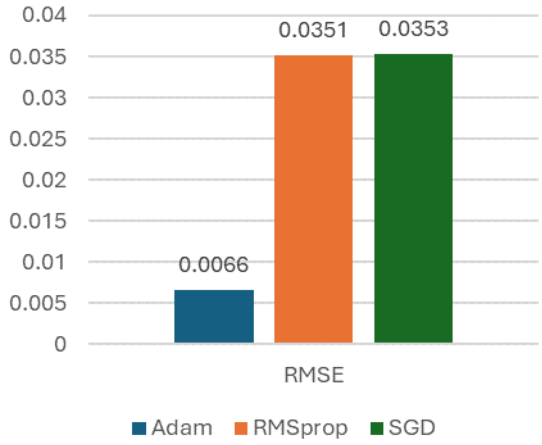

Figure 13. The influence of optimizer through experiments.

training intervals, 102 hours for validating intervals, and 408 hours for testing periods, respectively. The curve of the estimated voltage of the suggested method learned on FC1 correctly tracks the fluctuation of the voltage of the stack, providing crucial information about the voltage of the stack deterioration pattern in advance, as seen by FC1's estimation outcomes displayed in Figure 14. The curves of the estimated voltage of the suggested model trained on FC2 appear to properly correspond to the voltage of the stack fluctuation, based on the FC2's estimation outcomes displayed in Figure 15.

Table 3. Experimental Analysis of the influence of hyper-parameter on prediction results.

	Hyper-parameter	Value	RMSE
LSTM_1	Number of cells	64	0.0514
		128	**0.0113**
		256	0.0210
LSTM_2	Number of cells	64	0.0623
		128	**0.0113**
		256	0.0154
LSTM_3	Number of cells	**64**	**0.0113**
		128	0.0333
		256	0.0522
LSTM_4	Number of cells	**64**	**0.0113**
		128	0.0461
		256	0.0813
DENSE_1	Number of cells	8	0.0947
		16	0.0716
		32	**0.0113**
DENSE_2	Number of cells	**8**	**0.0113**
		16	0.0689
		32	0.0901
Dropout rate	Dropout rate value	0.3	0.143
		0.4	0.096
		0.5	**0.0113**
Le	Learning rate value	0.0001	0.125

		0.001	**0.0113**
		0.002	0.0162
		0.01	0.213
Optimizer	optimizers	**Adam**	**0.0113**
		RMSprop	0.164
		SGD	0.314

Table 4. The DA-LSTM hyperparameters.

Parameter	value
LSTM_1 number of cells	128
LSTM_2 number of cells	128
LSTM_3 number of cells	64
LSTM_4 number of cells	64
Dense_1 number of nodes	32
Dense_2 number of nodes	8
Dropout rate	0.5
Learning rate	0.001
Max no. of epoch	1000
Loss	MSE
Optimizer	Adam

5. Results and discussion

5.1 Comparison Results

In this specific section, the outcomes attained by DA-LSTM for the testing subset of the 2014 PHM Data Challenge Dataset are juxtaposed against rival models to exhibit their effectiveness and efficiency. These outcomes are measured using the RMSE metric to highlight how well the models reduce the gap between the expected and targeted RUL. The results of the model proposed in the 2014 PHM Data Challenge Dataset are contrasted with various competing models such as Fusion [48], 1 input-ESN [49], 2 input-ESN [49], 3 input-ESN [49], SAE-DNN [50], ML-DNN [50], and RCLMA [51]. The superior efficacy of the system is illustrated through the presentation of findings, as evidenced by the RMSE values detailed in Table 5.

The significant outcomes are highlighted in bold style. The presented table reveals that DA-LSTM outperforms all other models investigated in terms of the RMSE over the testing dataset, achieving RMSE values of 0.0066. Upon contrasting our results with the best outcomes attained by different referenced models, our suggested model showcases a decrease in RMSE of 60% for the 2014 PHM Data Challenge Dataset. This proposed model is considered a robust choice for tackling this problem as it has the capacity to operate most efficiently based on the RMSE measure, which assigns equal importance to predictions made earlier on and later. To visually exhibit the efficacy of the recommended model, Figure 16 has been included to display the RMSE values obtained from various algorithms for the testing set.

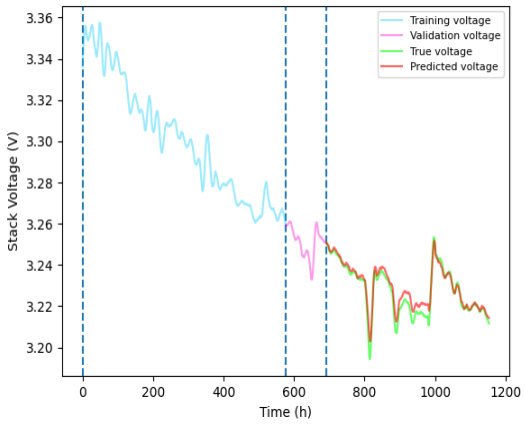

Figure 14. the curves of FC1's estimated stack voltage.

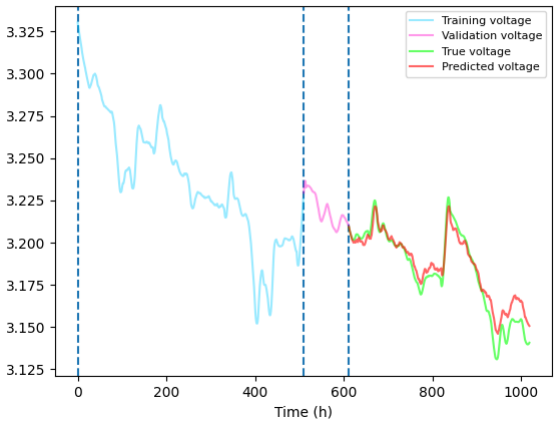

Figure 15. the curves of FC2's estimated stack voltage.

Table 5. Performance comparison with competing models for the testing set.

Models	**RMSE**
Fusion	0.0165
ESN based on 3 inputs	0.01804
ESN based on 2 inputs	0.02192
ESN based on 1 input	0.03354
SAE-DNN	0.2867
ML-DNN	0.1422
RCLMA	0.01785
Proposed method (DA-LSTM)	**0.0066**

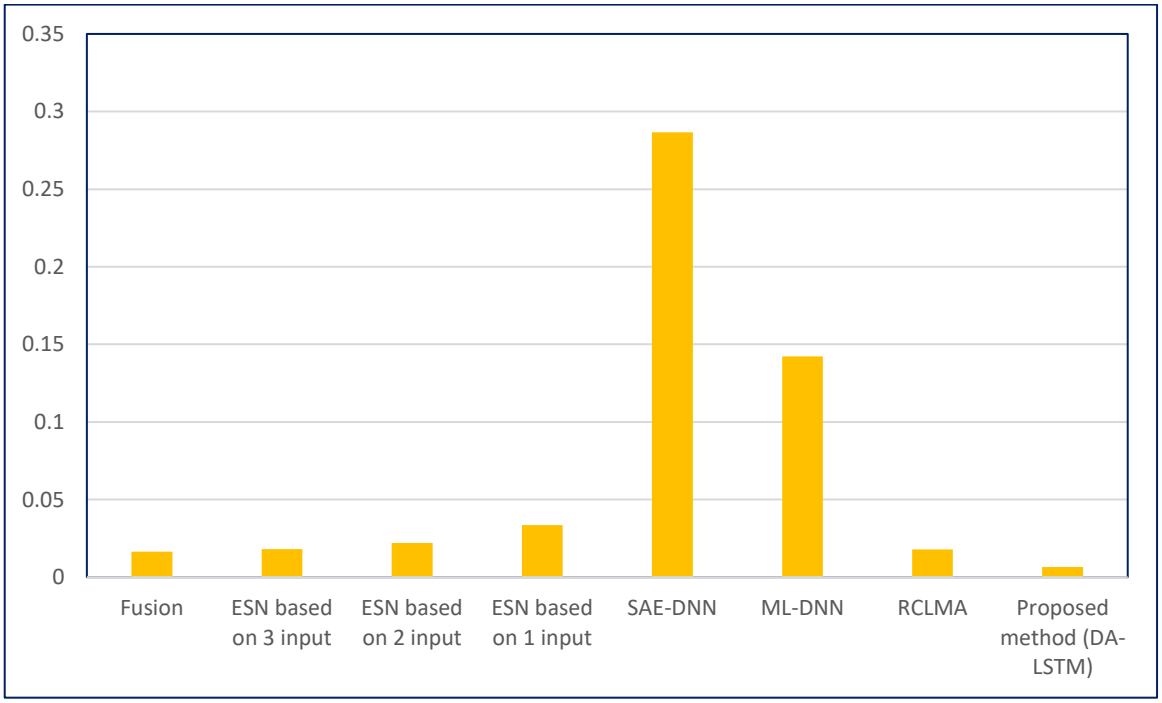

Figure 16. illustrates the representation of RMSE values acquired from different models.

5.2 Ablation Study Results

To assess the effects of each component in the DA-LSTM individually, ablation experiments are conducted using the proposed methodology. This section outlines four experiments that are conducted under the same conditions as the

DA-LSTM. The experiments involve the utilization of various neural network architectures, such as an LSTM model lacking an attention mechanism, a hybrid model combining LSTM and self-attention mechanism (referred to as Self-LSTM), a hybrid model combining LSTM and scaled dot-product attention mechanism (referred to as Scaled-LSTM), and a model combining LSTM with both self-attention mechanism and scaled dot-product attention method (DA-LSTM). Furthermore, the outcomes of the experiments are meticulously laid out in Table 6 and visually depicted in Figure 17. The ATCN-LSTM displays superior performance in the RMSE metric compared to LSTM, Self-LSTM, and Scaled-LSTM across the testing dataset. The RMSE value of the proposed approach DA-LSTM is registered at 0.0066, indicating enhancements of 81.3%, 77.5%, and 80.7% over LSTM, Self-LSTM, and Scaled-LSTM, respectively.

Table 6. The results of the ablation study.

Model	RMSE
LSTM	0.0354
Self-LSTM	0.0294
Scaled-LSTM	0.0311
Self-Scaled-LSTM (DA-LSTM)	**0.0066**

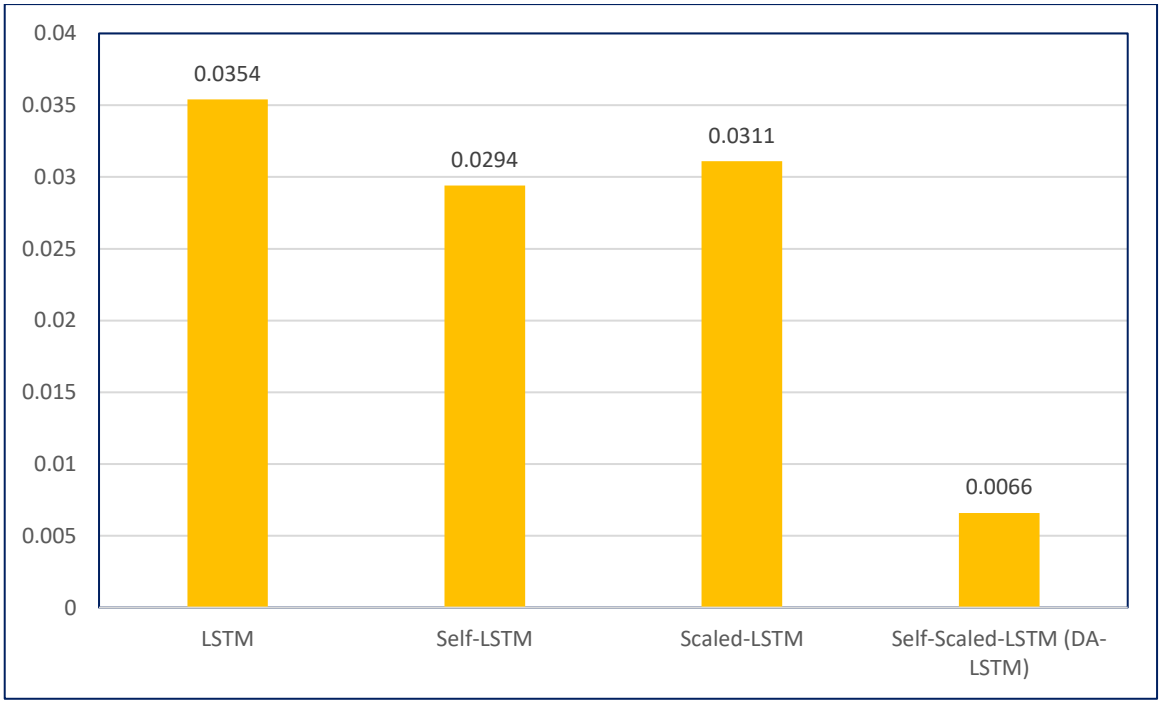

Figure 17. The representation of RMSE values acquired through ablation experiments.

6. Conclusion

This research study presents a new data-driven deep learning method called DA-LSTM, designed for the estimation of Proton Exchange Membrane Fuel Cells' (PEMFCs') Remaining Useful Life (RUL). The DA-LSTM model combines LSTM, a self-attention mechanism, and a scaled dot-product attention mechanism. Long sequences are captured using LSTM, which helps the model comprehend intricate temporal correlations in time-based data. Additionally, two attention mechanisms are integrated to synchronize input and output sequences by considering the context or significance of the input sequence. The self-attention mechanism is utilized to identify relationships among different time points, while the hidden state layer neurons' weights are set via the scaled dot-product attention mechanism. The RUL forecasts are generated through a Fully Connected (FC) network. The experimental assessment was carried out using the PHM 2014

Challenge dataset. Comparative analysis with established models in the domain illustrated that our proposed model achieved a 60% reduction in RMSE for the testing dataset.

Funding

This research was conducted without external funding support.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

References

[1] X. Sun, M. Xie, F. Zhou, X. Wu, J. Fu, and J. Liu, "Hierarchical evolutionary construction of neural network models for an Atkinson cycle engine with double injection strategy based on the PSO-Nadam algorithm," *Fuel*, vol. 333, p. 126531, Feb. 2023, doi: 10.1016/j.fuel.2022.126531.

[2] T. Cai, D. Zhao, and J. E, "Bluff-body effect on thermal and NOx emission characteristics in a micro-planar combustor fueled with premixed ammonia-oxygen," *Chem. Eng. Process. - Process Intensif.*, vol. 153, p. 107979, Jul. 2020, doi: 10.1016/j.cep.2020.107979.

[3] T. Cai and D. Zhao, "Enhancing and assessing ammonia-air combustion performance by blending with dimethyl ether," *Renew. Sustain. Energy Rev.*, vol. 156, p. 112003, Mar. 2022, doi: 10.1016/j.rser.2021.112003.

[4] S. Ni and D. Zhao, "NOx emission reduction in ammonia-powered micro-combustors by partially inserting porous medium under fuel-rich condition," *Chem. Eng. J.*, vol. 434, p. 134680, Apr. 2022, doi: 10.1016/j.cej.2022.134680.

[5] H. Dong, J. Fu, Z. Zhao, Q. Liu, Y. Li, and J. Liu, "A comparative study on the energy flow of a conventional gasoline-powered vehicle and a new dual clutch parallel-series plug-in hybrid electric vehicle under NEDC," *Energy Convers. Manag.*, vol. 218, p. 113019, Aug. 2020, doi: 10.1016/j.enconman.2020.113019.

[6] X. Sun, J. Fu, H. Yang, M. Xie, and J. Liu, "An energy management strategy for plug-in hybrid electric vehicles based on deep learning and improved model predictive control," *Energy*, vol. 269, p. 126772, Apr. 2023, doi: 10.1016/j.energy.2023.126772.

[7] A. Joshi, R. Sharma, and B. Baral, "Comparative life cycle assessment of conventional combustion engine vehicle, battery electric vehicle and fuel cell electric vehicle in Nepal," *J. Clean. Prod.*, vol. 379, p. 134407, Dec. 2022, doi: 10.1016/j.jclepro.2022.134407.

[8] Q. Zhigang, *Proton exchange membrane fuel cells*. 2013. [Online]. Available: https://doi.org/10.1201/b15499

[9] Y. Wu, E. Breaz, F. Gao, D. Paire, and A. Miraoui, "Nonlinear Performance Degradation Prediction of Proton Exchange Membrane Fuel Cells Using Relevance Vector Machine," *IEEE Trans. Energy Convers.*, vol. 31, no. 4, pp. 1570–1582, Dec. 2016, doi: 10.1109/TEC.2016.2582531.

[10] R. E. Silva *et al.*, "Proton exchange membrane fuel cell degradation prediction based on Adaptive Neuro-Fuzzy Inference Systems," *Int. J. Hydrog. Energy*, vol. 39, no. 21, pp. 11128–11144, Jul. 2014, doi: 10.1016/j.ijhydene.2014.05.005.

[11] S. Morando, S. Jemei, R. Gouriveau, N. Zerhouni, and D. Hissel, "Fuel Cells prognostics using echo state network," in *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, Nov. 2013, pp. 1632–1637. doi: 10.1109/IECON.2013.6699377.

[12] D. Zhou, Y. Wu, F. Gao, E. Breaz, A. Ravey, and A. Miraoui, "Degradation Prediction of PEM Fuel Cell Stack Based on Multiphysical Aging Model With Particle Filter Approach," *IEEE Trans. Ind. Appl.*, vol. 53, no. 4, pp. 4041–4052, Aug. 2017, doi: 10.1109/TIA.2017.2680406.

[13] Y. Wang, Y. Hu, and C. Sun, "Remaining Useful Life Prediction for Proton Exchange Membrane Fuel Cell Using Stochastic Fusion Filtering∗∗This work was supported by the National Natural Science Foundation of China under Grants 61773055 and 61520106009,the Fundamental Research Funds for the Central Universities of China under Grant FRF-TP-16-029A3.," *5th IFAC Workshop Min. Miner. Met. Process. MMM 2018*, vol. 51, no. 21, pp. 158–162, Jan. 2018, doi: 10.1016/j.ifacol.2018.09.409.

[14] X. Zhang, D. Yang, M. Luo, and Z. Dong, "Load profile based empirical model for the lifetime prediction of an automotive PEM fuel cell," *Int. J. Hydrog. Energy*, vol. 42, no. 16, pp. 11868–11878, Apr. 2017, doi: 10.1016/j.ijhydene.2017.02.146.

[15] M. Mayur, M. Gerard, P. Schott, and W. G. Bessler, "Lifetime Prediction of a Polymer Electrolyte Membrane Fuel Cell under Automotive Load Cycling Using a Physically-Based Catalyst Degradation Model," *Energies*, vol. 11, no. 8, 2018, doi: 10.3390/en11082054.

[16] Z. Hu, L. Xu, J. Li, M. Ouyang, Z. Song, and H. Huang, "A reconstructed fuel cell life-prediction model for a fuel cell hybrid city bus," *Energy Convers. Manag.*, vol. 156, pp. 723–732, Jan. 2018, doi: 10.1016/j.enconman.2017.11.069.

[17] L. Zhu and J. Chen, "Prognostics of PEM fuel cells based on Gaussian process state space models," *Energy*, vol. 149, pp. 63–73, Apr. 2018, doi: 10.1016/j.energy.2018.02.016.

[18] H. Chen, P. Pei, and M. Song, "Lifetime prediction and the economic lifetime of Proton Exchange Membrane fuel cells," *Appl. Energy*, vol. 142, pp. 154–163, Mar. 2015, doi: 10.1016/j.apenergy.2014.12.062.

[19] K. Chen, S. Laghrouche, and A. Djerdir, "Fuel cell health prognosis using Unscented Kalman Filter: Postal fuel cell electric vehicles case study," *Int. J. Hydrog. Energy*, vol. 44, no. 3, pp. 1930–1939, Jan. 2019, doi: 10.1016/j.ijhydene.2018.11.100.
[20] M. Ou *et al.*, "A novel approach based on semi-empirical model for degradation prediction of fuel cells," *J. Power Sources*, vol. 488, p. 229435, Mar. 2021, doi: 10.1016/j.jpowsour.2020.229435.
[21] E. Lechartier, E. Laffly, M.-C. Péra, R. Gouriveau, D. Hissel, and N. Zerhouni, "Proton exchange membrane fuel cell behavioral model suitable for prognostics," *Int. J. Hydrog. Energy*, vol. 40, no. 26, pp. 8384–8397, Jul. 2015, doi: 10.1016/j.ijhydene.2015.04.099.
[22] J. Wu, K. Hu, Y. Cheng, H. Zhu, X. Shao, and Y. Wang, "Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network," *ISA Trans.*, vol. 97, pp. 241–250, Feb. 2020, doi: 10.1016/j.isatra.2019.07.004.
[23] S. Fu, Y. Zhang, L. Lin, M. Zhao, and S. Zhong, "Deep residual LSTM with domain-invariance for remaining useful life prediction across domains," *Reliab. Eng. Syst. Saf.*, vol. 216, p. 108012, Dec. 2021, doi: 10.1016/j.ress.2021.108012.
[24] W. Zhang, D. Yang, and H. Wang, "Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey," *IEEE Syst. J.*, vol. 13, no. 3, pp. 2213–2227, Sep. 2019, doi: 10.1109/JSYST.2019.2905565.
[25] Z. Gao, C. Cecati, and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015, doi: 10.1109/TIE.2015.2417501.
[26] M. Yan, X. Wang, B. Wang, M. Chang, and I. Muhammad, "Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model," *ISA Trans.*, vol. 98, pp. 471–482, Mar. 2020, doi: 10.1016/j.isatra.2019.08.058.
[27] J. Ben Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, and F. Fnaiech, "Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network," *Mech. Syst. Signal Process.*, vol. 56–57, pp. 150–172, May 2015, doi: 10.1016/j.ymssp.2014.10.014.
[28] F. Shen and R. Yan, "A New Intermediate-Domain SVM-Based Transfer Model for Rolling Bearing RUL Prediction," *IEEEASME Trans. Mechatron.*, vol. 27, no. 3, pp. 1357–1369, Jun. 2022, doi: 10.1109/TMECH.2021.3094986.
[29] J. Nan, B. Deng, W. Cao, and Z. Tan, "Prediction for the Remaining Useful Life of Lithium–Ion Battery Based on RVM-GM with Dynamic Size of Moving Window," *World Electr. Veh. J.*, vol. 13, no. 2, 2022, doi: 10.3390/wevj13020025.
[30] C. Zhang, Y. He, L. Yuan, S. Xiang, and J. Wang, "Prognostics of Lithium-Ion Batteries Based on Wavelet Denoising and DE-RVM," *Comput. Intell. Neurosci.*, vol. 2015, p. 918305, Aug. 2015, doi: 10.1155/2015/918305.
[31] J. Tang *et al.*, "Rolling bearing remaining useful life prediction via weight tracking relevance vector machine," *Meas. Sci. Technol.*, vol. 32, no. 2, p. 024006, Dec. 2020, doi: 10.1088/1361-6501/abbe3b.
[32] J. -Y. Wu, M. Wu, Z. Chen, X. -L. Li, and R. Yan, "Degradation-Aware Remaining Useful Life Prediction With LSTM Autoencoder," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021, doi: 10.1109/TIM.2021.3055788.
[33] X. Wang, T. Wang, A. Ming, W. Zhang, A. Li, and F. Chu, "Spatiotemporal non-negative projected convolutional network with bidirectional NMF and 3DCNN for remaining useful life estimation of bearings," *Neurocomputing*, vol. 450, pp. 294–310, Aug. 2021, doi: 10.1016/j.neucom.2021.04.048.
[34] H.-K. Wang, Y. Cheng, and K. Song, "Remaining Useful Life Estimation of Aircraft Engines Using a Joint Deep Learning Model Based on TCNN and Transformer," *Comput. Intell. Neurosci.*, vol. 2021, p. 5185938, Nov. 2021, doi: 10.1155/2021/5185938.
[35] J. Zhu, N. Chen, and W. Peng, "Estimation of Bearing Remaining Useful Life Based on Multiscale Convolutional Neural Network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3208–3216, Apr. 2019, doi: 10.1109/TIE.2018.2844856.
[36] R. Ma, T. Yang, E. Breaz, Z. Li, P. Briois, and F. Gao, "Data-driven proton exchange membrane fuel cell degradation predication through deep learning method," *Appl. Energy*, vol. 231, pp. 102–115, Dec. 2018, doi: 10.1016/j.apenergy.2018.09.111.
[37] R. Mezzi, N. Yousfi-Steiner, M. C. Péra, D. Hissel, and L. Larger, "An Echo State Network for fuel cell lifetime prediction under a dynamic micro-cogeneration load profile," *Appl. Energy*, vol. 283, p. 116297, Feb. 2021, doi: 10.1016/j.apenergy.2020.116297.
[38] F.-K. Wang, Z. E. Amogne, J.-H. Chou, and C. Tseng, "Online remaining useful life prediction of lithium-ion batteries using bidirectional long short-term memory with attention mechanism," *Energy*, vol. 254, p. 124344, Sep. 2022, doi: 10.1016/j.energy.2022.124344.
[39] J. Chen, H. Jing, Y. Chang, and Q. Liu, "Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process," *Reliab. Eng. Syst. Saf.*, vol. 185, pp. 372–382, May 2019, doi: 10.1016/j.ress.2019.01.006.
[40] Y. Cheng, N. Zerhouni, and C. Lu, "A hybrid remaining useful life prognostic method for proton exchange membrane fuel cell," *Int. J. Hydrog. Energy*, vol. 43, no. 27, pp. 12314–12327, Jul. 2018, doi: 10.1016/j.ijhydene.2018.04.160.
[41] H. Liu, J. Chen, D. Hissel, and H. Su, "Remaining useful life estimation for proton exchange membrane fuel cells using a hybrid method," *Appl. Energy*, vol. 237, pp. 910–919, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.023.
[42] X. Li, Y. Ma, and J. Zhu, "An online dual filters RUL prediction method of lithium-ion battery based on unscented particle filter and least squares support vector machine," *Measurement*, vol. 184, p. 109935, Nov. 2021, doi: 10.1016/j.measurement.2021.109935.
[43] R. Gouriveau *et al.*, "IEEE PHM 2014 data challenge: Outline, experiments, scoring of results, winners," presented at the Proc. IEEE Conf. Prognostics Health Manage, 2014, pp. 1–6.
[44] Y. Zhao, Y. Li, Y. Cao, L. Jiang, J. Wan, and C. Rehtanz, "An RNN With Small Sequence Trained by Multi-Level Optimization for SOC Estimation in Li-Ion Battery Applications," *IEEE Trans. Veh. Technol.*, vol. 72, no. 9, pp. 11469–11481, Sep. 2023, doi: 10.1109/TVT.2023.3267500.

[45] Z. Li, R. Outbib, S. Giurgea, and D. Hissel, "Diagnosis for PEMFC Systems: A Data-Driven Approach With the Capabilities of Online Adaptation and Novel Fault Detection," *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 5164–5174, Aug. 2015, doi: 10.1109/TIE.2015.2418324.
[46] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *ArXiv Prepr. ArXiv14126980*, 2014.
[48] R. Ma *et al.*, "Data-Fusion Prognostics of Proton Exchange Membrane Fuel Cell Degradation," *IEEE Trans. Ind. Appl.*, vol. 55, no. 4, pp. 4321–4331, Aug. 2019, doi: 10.1109/TIA.2019.2911846.
[49] Z. Hua, Z. Zheng, M.-C. Péra, and F. Gao, "Remaining useful life prediction of PEMFC systems based on the multi-input echo state network," *Appl. Energy*, vol. 265, p. 114791, May 2020, doi: 10.1016/j.apenergy.2020.114791.
[50] F. -K. Wang, Z. E. Amogne, and J. -H. Chou, "A Hybrid Method for Remaining Useful Life Prediction of Proton Exchange Membrane Fuel Cell Stack," *IEEE Access*, vol. 9, pp. 40486–40495, 2021, doi: 10.1109/ACCESS.2021.3064684.
[51] X. Sun, M. Xie, J. Fu, F. Zhou, and J. Liu, "An improved neural network model for predicting the remaining useful life of proton exchange membrane fuel cells," *Int. J. Hydrog. Energy*, vol. 48, no. 65, pp. 25499–25511, Jul. 2023, doi: 10.1016/j.ijhydene.2023.03.219.