Eur. Phys. J. Spec. Top.
https://doi.org/10.1140/epjs/s11734-024-01408-8

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Regular Article

Prediction of PEM fuel cell performance degradation using bidirectional long short-term memory with chimp optimization algorithm

Başak Ekinci[1,a], İlker Dursun[1], Zeynep Garip[2], and Ekin Ekinci[2]

[1] Electrical and Electronics Engineering Department, Faculty of Technology, Sakarya University of Applied Sciences, Sakarya, Turkey
[2] Computer Engineering Department, Faculty of Technology, Sakarya University of Applied Sciences, Sakarya, Turkey

Received 1 October 2024 / Accepted 14 November 2024

Abstract The proton exchange membrane fuel cells (PEMFC) are among the most promising technologies for efficiently converting hydrogen into electricity with minimal emissions. Significant advancements have been made in enhancing the performance, durability, and cost-effectiveness of PEMFC. However, these cells still face challenges related to performance degradation over time. Therefore, this study focuses on voltage prediction, which is one of the most important key factors for assessing fuel cell performance and extending its lifetime. This study combines the chimpanzee optimization algorithm (ChOA) with long short-term memory (LSTM), stacked LSTM, and bidirectional LSTM (BiLSTM) networks to predict performance degradation in PEM fuel cells. Initially, features from the PEMFC time-series data are reduced using the ChOA to select the most informative ones. These selected features are subsequently input into the corresponding LSTM networks to enhance the accuracy of PEMFC performance degradation predictions. The experimental results in terms of root mean squared error (RMSE) indicate that the ChOA variants—specifically, ChOALSTM, ChOAStackedLSTM, and ChOABiLSTM—achieved prediction accuracies of 0.012, 0.014, and 0.007 on the IEEE PHM 2014 DATA Challenge dataset, respectively. The comparative and statistical results obtained from the proposed ChOABiLSTM model demonstrate its superior accuracy and robustness compared to its variants and other state-of-the-art algorithms.

1 Introduction

Due to its high energy conversion efficiency, low operating temperature, and environmental friendly by products, PEMFCs are considered potential power generation system for various applications [33]. PEMFCs have garnered significant attention for portable, stationary power generation and transportation applications [28]. However, performance degradation in PEMFCs presents substantial challenges for effective implementation, particularly in applications requiring high reliability, extended lifespans, and low maintenance costs. Addressing these degradation mechanisms can improve the durability and cost-effectiveness of PEMFCs, thereby enhancing their viability for broader use in sustainable energy applications, including zero-emission vehicles and reliable backup power systems [5].

Prognostics and health management (PHM) plays a crucial role in predicting the aging trend by analyzing historical data and current operating conditions of PEMFC, aiming to minimize maintenance expenses and extend service life [35]. PHM approaches primarily consist of three categories: model-based methods, data-driven methods, and hybrid methods that combine both models and data [6]. The model-based approach integrates mathematical equations depicting the degradation phenomena in PEMFCs combined with state estimation techniques, enabling the estimation of State of Health (SOH) and prediction of remaining useful life [36]. Conversely, the data-driven approach represents a model-free strategy, utilizing machine learning (ML) frameworks to extract insights from large datasets gathered during PEMFC aging tests [14]. The model-data hybrid method combines both of the model-based approach and the data-driven method. Due to the difficulties of estimating the deterioration from the

[a] e-mail: 22500409003@subu.edu.tr (corresponding author)

Published online: 02 December 2024

physical characteristics and the recent rapid developments of ML techniques, data-driven methods have become a viable tool for predicting fuel cell performance. Deep learning (DL) methods outperform other ML algorithms in scalability and generalization, especially when processing large and complex data sets [31]. Consequently, DL-based approaches for predicting PEMFC performance degradation have gained increased interest, such as has drawn more interest including: LSTM [2, 15], gated recurrent unit (GRU) [19], stacked LSTM [32], BiLSTM [3, 20], LSTM-recurrent neural network (RNN) method [25], and echo state network (ESN) [17, 37]. Considerable research has focused on exploring various optimization techniques to enhance the accuracy of performance degradation predictions in PEMFCs. Table 1 summarizes studies on artificial intelligence algorithms applied to the same dataset in recent years.

Chimp optimization algorithm (ChOA) is a new metaheuristic algorithm that has not yet been applied to feature selection issues in PEMFC. Therefore, its potential effectiveness in addressing this problem has remained unexplored. As mentioned, the metaheuristic algorithm excels at solving the feature selection problem. It is highly effective at discovering better solutions for complex optimization problems compared to others [29]. Consequently, it is necessary to assess the performance of the proposed ChOA and demonstrate its effectiveness.

In this study, we propose a method based on the ChOA and BiLSTM to minimize high dimensionality and improve the precision of degradation prediction for PEMFC. ChOA selects the most informative features from the IEEE Prognostics and Health Management (PHM) 2014 Data Challenge dataset [26], significantly enhancing the degradation prediction accuracy of BiLSTM in the proposed model.

The main contributions are as follows:

- An approach for predicting PEMFC degradation is presented and verified using experimental data.
- ChOA as an encouraging feature selection method has been first proposed for PEMFC degradation prediction, applied with BiLSTM architecture. The prediction performance is compared with ML and LSTM models, their ChOA variants, and ranking-based feature selection algorithms.
- The result proves that the ChOABiLSTM model is superior in PEMFC degradation prediction.

The following paper is organized as: Sect. 2 describes the experimental PEMFC dataset and BiLSTM and ChOA algorithms. Section 3 describes the experimental setup, performance metrics, compares the prediction results with other algorithms, and discusses the results of prediction. Finally, Sect. 4 concludes the paper.

2 Materials and methods

2.1 Experimental dataset

In this study, provided by IEEE PHM 2014 Data Challenge [26] for the purpose of estimating the RUL of PEMFC is used as experimental dataset. FC2 dataset which is collected under rippled current conditions designated as the training and testing data. For FC2, the initial 505 h of data were utilized for training purposes (65,535 samples), while the subsequent data from 505 to 1020 h (61,834 samples) were designated for testing.

In FC2, the proton-exchange-membrane fuel cell underwent testing for 1020 h, maintaining a constant current of 70 A and a 10% fluctuation occurring at a frequency of 5 kHz throughout the experiment. The PEMFC stack employed in the experiment comprises 5 cells, namely the BZ 100 model produced by Ulmer. Each cell possesses an active area measuring 100 cm^2, with a nominal current density of 0.7 A/cm^2 and a maximum current density of 1 A/cm^2. The total stack is capable of producing a maximum power output of 600 W. The fuel cell stack operated at a temperature of 54 °C. The relative pressure at the anode and cathode inlet was maintained at 1.3 bar, with the air relative humidity of the gas set at 52% and sampling interval for every 30 s.

Various aging parameters are assessed through sensors positioned at both the inlet and outlet of the PEMFC stack to calculate stack voltage Utot (V). These parameters are detailed in Table 2, offering valuable insights into the stack's condition and performance.

2.2 BiLSTM

Bidirectional LSTM (BiLSTM) first introduced by Graves and Schmidhuber [13] is an expansion of traditional LSTM architecture. Unlike unidirectional LSTM, which processes data only in one direction (backward propagation), BiLSTM integrates both forward and backward propagation to extract information from both past and future data sequences. This dual-directional processing enables the network to make more informed predictions by considering a broader context. In BiLSTM, historical information complements the current data, creating a richer dependency chain for better understanding of the input sequences. Figure 1 illustrates the structure of BiLSTM.

Table 1 Relevant literature and their findings

Ref./year	Features	Methods	Error metrics	Results
[28]	H2 inlet temperature, air outlet pressure, and H2 outlet flow rate	GA-NARX	IAE, ISE, ITAE, ITSE, MSE, R^2	The experimental findings show that, in terms of prediction accuracy, the GA-NARX method outperforms GA-BPNN and GA-TDNN techniques
[38]	Air outlet flow, H2 inlet temperature, current, cooling water flow, air outlet pressure, H2 outlet temperature, cooling water outlet temperature	B-GRU	MSE, RMSE, MAE	The suggested B-GRU performs better in datasets based on point and interval estimate findings
[3]	All	Bayesian-Bi-LSTM	RMSE, MAPE, Max AE	Performance deterioration prediction can be accomplished with greater accuracy using the suggested Bayesian optimization strategy
[1]	All	Dil-CNN-att	RMSE, MAPE,	The suggested method may forecast not only the degradation tendency but also the dynamics of the degradation behavior, according to the results
[4]	The historical status, load current, relative humidity, temperature, and hydrogen pressure	Wavelet analysis and NARX	MSE, RE	Using this new prognostic technique, which encompasses PEMFC degradation over a broad range of operating circumstances, a robust model is produced
[33]	All	DBN-ELM	RMSE, MAPE, R^2	When compared to the current traditional methods, the suggested strategy performs better in predictions, regardless of the training phase or the number of steps ahead in the prediction

Table 1 (continued)

Ref./year	Features	Methods	Error metrics	Results
[7]	All	Stacked ESN-GA	RMSE, MAPE	The accuracy and generalization performance of the suggested method are better than those of traditional prediction methods
[34]	All	SSA-DGP	RMSE, MAE, RE	The case study demonstrates the suggested method's accuracy and dependability
[24]	All	Bi-LSTM-GRU and ESN	RMSE, MAE, R^2	The prediction accuracy has increased with respect to the conventional model
[30]	All	CEEMD-CNN-LSTM	MAE, MAPE, RMSE	The proposed model enhance the performance of the voltage deterioration prediction
[5]	All	PSO-GNNM	MPE	The suggested technique has the ability to accurately predict PEMFC degradation for a variety of applications
[32]	All	S-LSTM	RMSE MAPE	In terms of RMSE and MAPE, the suggested model outperforms the traditional models
[6]	All	GA-ELM	APE, MPE	Compared to other conventional data-based prognostics methods, the suggested method delivers improved voltage degradation prediction for PEMFC under dynamic load current
In this study	U3, U4, U5, I, ToutH2, PinAIR, DinH2, DinAIR and Utot	ChOABiLSTM	RMSE, MAE, R^2	The experimental results confirm that the proposed ChOABiLSTM model is superior in predicting PEMFC degradation

Using the LSTM units in two directions, one can enhance the model's accuracy by improving its capacity to learn long-term dependencies [8].

To validate the non-cyclic expansion graph, there are no available data forthe flow between the forward and backward layers [12]. Formulas for the forward LSTM are given below [10]

$$i_t^{(f)} = \sigma\left(W_i^{(f)} \cdot x_t + U_i^{(f)} \cdot h_{t-1}^{(f)} + b_i^{(f)}\right) \tag{1}$$

Table 2 The details of features of IEEE PHM 2014 Data Challenge dataset

Feature	Description
Time	Aging time (h)
U1 to U5	Voltage of single cells (V)
I; J	Current (A) and current density (A/cm^2)
TinH2; ToutH2	Inlet and outlet temperatures of H2 (^{o}C)
TinAIR; ToutAIR	Inlet and outlet temperatures of air (^{o}C)
TinWAT; ToutWAT	Inlet and outlet temp. of cooling water (^{o}C)
PinH2; PoutH2	Inlet and outlet pressure of H2 (mbara)
PinAIR; PoutAIR	Inlet and outlet pressure of air (mbara)
DinH2; DoutH2	Inlet and outlet flow rate of H2 (l/mn)
DinAIR; DoutAIR	Inlet and outlet flow rate of air (l/mn)
DWAT	Flow rate of cooling water (l/mn)
HrAIRFC	Inlet hygrometry (air)—estimated (%)
Utot	Stack voltage (V)

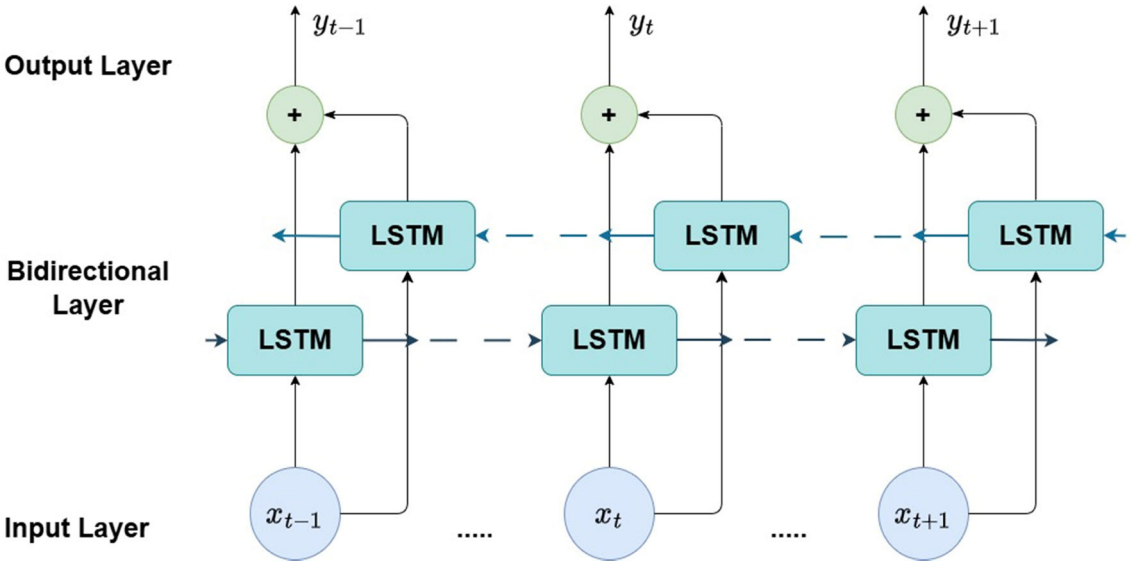


Fig. 1 BiLSTM architecture [18]

$$f_t^{(f)} = \sigma\left(W_f^{(f)} \cdot x_t + U_f^{(f)} \cdot h_{t-1}^{(f)} + b_f^{(f)}\right) \tag{2}$$

$$\widetilde{C}_t^{(f)} = \tanh\left(W_C^{(f)} \cdot x_t + U_C^{(f)} \cdot h_{t-1}^{(f)} + b_C^{(f)}\right) \tag{3}$$

$$C_t^{(f)} = f_t \cdot C_{t-1}^{(f)} + i_t^{(f)} \cdot \widetilde{C}_t^{(f)} \tag{4}$$

$$h_t^{(f)} = o_t^{(f)} \cdot \tanh\left(C_t^{(f)}\right). \tag{5}$$

Formulas for the backward LSTM are given below [10]

$$i_t^{(b)} = \sigma\left(W_i^{(b)} \cdot x_t + U_i^{(b)} \cdot h_{t-1}^{(b)} + b_i^{(b)}\right) \tag{6}$$

$$f_t^{(b)} = \sigma\left(W_f^{(b)} \cdot x_t + U_f^{(b)} \cdot h_{t-1}^{(b)} + b_f^{(b)}\right) \tag{7}$$

$$\widetilde{C}_t^{(b)} = \tanh\left(W_C^{(b)} \cdot x_t + U_C^{(b)} \cdot h_{t-1}^{(b)} + b_C^{(b)}\right) \tag{8}$$

$$C_t^{(b)} = f_t \cdot C_{t-1}^{(b)} + i_t^{(b)} \cdot \widetilde{C}_t^{(b)} \tag{9}$$

$$h_t^{(b)} = o_t^{(b)} \cdot \tanh\left(C_t^{(b)}\right). \tag{10}$$

Concatenating the outputs from the forward and backward LSTMs results in the final output at time step t

$$h_t = \left[h_t^{(f)}, h_t^{(b)}\right]. \tag{11}$$

In the formulas above, x_t is the input vector for time t. While $h_t^{(f)}$ is the forward LSTM's hidden state output, $h_t^{(b)}$ is the backward LSTM's hidden state output at time step t. $C_t^{(f)}$ and $C_t^{(b)}$ are the cell states of the forward and backward LSTM at time step t, respectively. W and U are weight matrices and b is the bias.

2.3 Chimp optimization algorithm (ChOA)

The ChOA is a metaheuristic algorithm that simulates the hunting behavior of chimpanzees [23]. Chimpanzee communities are classified into four groups by the ChOA algorithm: driver, barrier, attacker, and chaser [21]. Drivers are tasked with tracking the prey without capturing it, whereas chasers are responsible for swiftly pursuing and hunting down the prey. Barriers are responsible for limiting the movement of prey. The population leader is the attacker among them and predicts the prey's escape routes, bringing it back to the pursuit area [27]. There are phases of exploration and exploitation in the hunting process. Exploration consists of driving, blocking, and chasing the prey and exploitation which consists of attacking the prey [23].

In the exploration phase, the first two roles of the team hunting, driver and chase, are mathematically modeled by Eqs. 12 and 13. X_{prey} is a vector containing a prey location, X_{chimp} is a vector featuring a chimp position, and t is the number of iterations that are currently in progress

$$d = |n_{\text{vector}}.X_{\text{chimp}}(t) - n_{\text{vector}}.X_{\text{prey}}(t)| \tag{12}$$

$$X_{\text{chimp}}(t+1) = X_{\text{prey}}(t) - s_{\text{vector}}.d. \tag{13}$$

s_{vector}, v_{vector}, and n_{vector} stand for the coefficient vectors and are calculated by Eqs. 14–16. rand_1 and rand_2 indicate random integers in the interval [0, 1], while d_{nl} is a parameter whose value declined non-linearly from 2.5 to 0 during the iteration process. A vector called $\text{Chaotic}_{\text{value}}$ is the result of several chaotic mappings

$$s_{\text{vector}} = (2 \times d_{nl} \times \text{rand}_1) - d_{nl} \tag{14}$$

$$v_{\text{vector}} = 2 \times \text{rand}_2 \tag{15}$$

$$n_{\text{vector}} = \text{Chaotic}_{\text{value}}. \tag{16}$$

In the exploitation phase (attacker), through chasing, blocking, and driving, the chimp uncovers and encircles the prey. While all chimps participate in hunting, attackers execute the final assault. The initial positions of the first attacker, chaser, barrier, and driver are regarded as the primary optimal locations for the prey [16, 22]. As a result, the best chimpanzees are -updated according to their location using Eqs. 17–19

$$\begin{cases} d_{\text{Driver}} = |C_1 X_{\text{Driver}} - m_1 X|, \\ d_{\text{Chaser}} = |C_2 X_{\text{Chaser}} - m_2 X|, \\ d_{\text{Driver}} = |C_3 X_{\text{Barrier}} - m_3 X|, \\ d_{\text{Driver}} = |C_4 X_{\text{Attacker}} - m_4 X|. \end{cases} \tag{17}$$

$$\begin{cases} X_1 = X_{\text{Driver}} - S_{\text{vector1}}(d_{\text{Driver}}), \\ X_2 = X_{\text{Chaser}} - S_{\text{vector2}}(d_{\text{Chaser}}), \\ X_3 = X_{\text{Barrier}} - S_{\text{vector3}}(d_{\text{Barrier}}) \\ X_4 = X_{\text{Attacker}} - S_{\text{vector4}}(d_{\text{Attacker}}). \end{cases} \tag{18}$$

$$X(t+1) = \frac{X_1 + X_2 + X_3 + X_4}{4}. \tag{19}$$

The m vector represents the chaotic behavior of chimps during the latter stage of the hunt, aimed at securing greater quantities of meat, leading to increased social benefits such as grooming or mating opportunities. It is

used to mimic the social motivation behavior of ChOA with Eq. 20. μ in the equation represents a random value varying in the range [0, 1]

$$X_{\text{Chimp}}(t+1) = \begin{cases} X_{\text{prey}}(t) - s_{\text{vector}}.\text{dif}\mu < 0.5, \\ \text{Chaotic}_{\text{value}}\text{if}\mu > 0.5. \end{cases} \tag{20}$$

Algorithm 1 Pseudo-code of ChOA

```
Initialize the chimp population x_i (i=1, 2,…,n)
Initialize f, m, a and c
Calculate the position of each chimp
Divide chimps randomly into independent groups
Until stopping condition is satisfied
Calculate the fitness of each chimp
X_Attacker=the best search agent
X_Chaser=the second best search agent
X_Barrier=the third the best search agent
X_Driver=the fourth the best search agent
while (t<maximum number of iterations)
    for each chimp:
        Extract the chimp's group
        Use its group strategy to update f, m and c
        Use f, m and c to calculate a and then d
    end for
        for each search chimp
            if(μ<0.5)
                if(|a|<1)
        Update the position of the current search agent by the Eq. (13)
                else if(|a|<1)
                    Select a random search agent
                end if
            if(μ>0.5)
        Update the position of the current search agent by the Eq. (20)
            end if
        end for
 Update f, m, a and c
 Update X_Attacker, X_Chaser, X_Barrier and X_Driver
 t=t+1
end while
return X_Attacker
```

2.4 The proposed ChOABiLSTM model

In this study, a new hybrid model is proposed by combining ChOA and BiLSTM models. A comprehensive evaluation of the prediction performance is provided with this hybrid model. In addition, the most suitable model for predicting PEM fuel cell performance degradation is determined. The ChOA model enhances prediction performance by selecting the most relevant and valuable features from the dataset, while the BiLSTM model makes predictions based on these selected features. BiLSTM processes the series in both forward and backward directions, captures contextual information and dependencies from past and future contexts, and encodes the time-series into the hidden state sequence. It successfully mitigates overfitting during the model training process. To improve the model's ability to generalize, the dropout layer which randomly removes neurons from the network with a predetermined probability during the training phase is used. This keeps the model from becoming overly dependent on any one neuron. At the end, encoded time-series data decoded in dense layer, and thus, the final output is produced. Figure 2 shows the proposed ChOABiLSTM prediction model.

3 Experimental study

3.1 Experimental setup

The ChOA feature selection and BiLSTM model implementation for stack voltage prediction are conducted using MATLAB R2022a and Python 3.10. MATLAB is employed for feature selection due to its robust computational capabilities, while Python (specifically the Keras library in Google Colab) is used to implement the BiLSTM

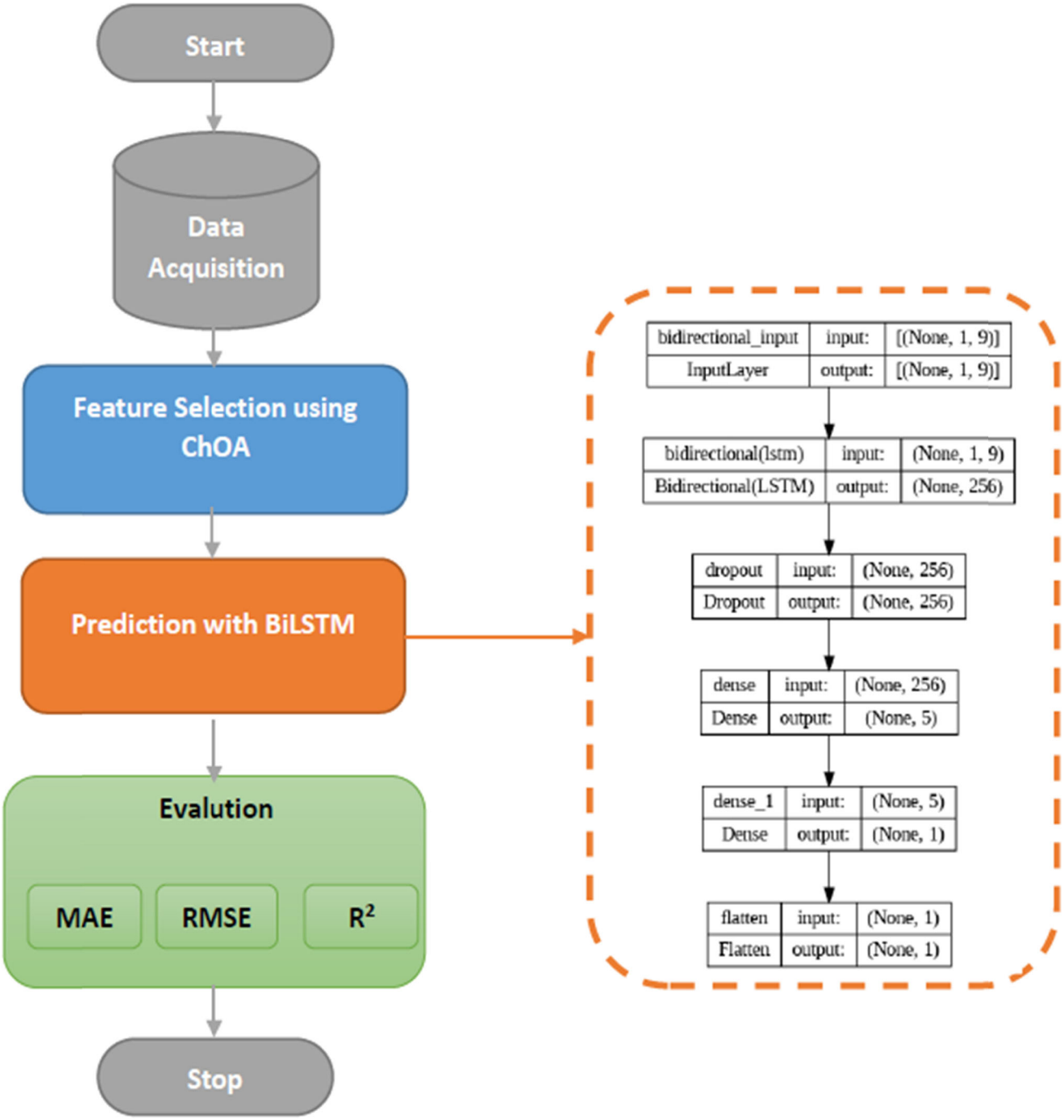


Fig. 2 The proposed ChOABiLSTM model

model. The system configuration for the experiments includes a GPU Tesla K80 with 12 GB of GDDR5 VRAM, an Intel Xeon Processor with two 2.20-GHz cores, and 13 GB RAM.

The model parameters must be chosen in accordance with the dataset to make the models useful. In DL models, it is essential to define the number of neurons in the input layer, the architecture and count of hidden layers, the number of neurons per hidden layer, the configuration of the fully connected layer, and the number of neurons in the output layer. Additionally, determining the optimal number of epochs and batch size is crucial for effective model training. The number of neurons in the input layer is typically determined by the number of features present in the input dataset.

At first, there are 24 neurons in the input layer, since the number of neurons in this layer is depend the number of features in the original dataset. After feature selection, the number of neurons is decreased; therefore, the

number of input neurons in the new network is also decreased. To calculate the size of the hidden neurons, there is no formula. The network consists of a fully connected layer with one neuron and a bidirectional layer (2 hidden layers) with 128 hidden neurons. The bidirectional layer is followed by dropout layer with 0.2 rate. After dropout layer, two dense layers with 5 and 1 neurons are added, respectively. A flatten layer is added, followed by a dense layer. Since $(t + 1)$ time aging is predicted using the 30-s prediction, the output layer's size with a single neuron is 1. Another crucial factor is the number of epochs and the epoch size is selected as 50. The batch size in our experiment is set to 72. For the purpose of creating DL models, optimizing the loss function is just as important as choosing the parameters. Recently, there has been a significant emphasis on the optimization of the loss function. We employed the Adam loss function optimizer during the model's training phase. Adam is a stochastic gradient descent algorithm that uses an exponential moving average to compute first- and second-order moments. Adam offers quick convergence in this regard, which is also the primary reason this study uses it. To determine the DL model's performance, an activation function must be used. Tanh is employed as the function of activation.

3.2 Performance metrics

In our experiments, we use the RMSE, the MAE, the R^2, and the loss value as error metrics to evaluate the performance of the experimental algorithms.

The RMSE is an estimate of the standard deviation of the residuals. It is the square root of the mean squared difference between the observed and predicted values. A higher RMSE indicates a poorer fit, with greater dispersion around the mean. A lower RMSE indicates a better fit, with smaller discrepancies between the observed and predicted values. It is therefore preferable to have a lower RMSE value. The formula for calculating the RMSE is as follows [10]:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}, \tag{21}$$

where the value of the ith observation is projected to be $\widehat{y_i}$, while the expected value is y_i. n is the total number of test observations.

The MAE is the average absolute deviation between the observed and predicted value. A lower MAE indicates that the prediction error of the model is lower. This corresponds to more accurate predictions. The MAE is calculated using the following formula [10]:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y_i}|. \tag{22}$$

R^2 measures how well the model explains the variance in the data and is the square of the correlation coefficient between the observed and predicted values. A high R^2 value shows a strong correlation between observations and predictions, while a low value shows a weak correlation. R^2 values lie between [0, 1]. When it is greater than 0.8, the model is considered acceptable; when it is lower, the model is considered less acceptable [9, 11]. R^2 is calculated according to the formula below [10]

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}, \tag{23}$$

where $\overline{y}$ is the mean of observed values (y).

3.3 Case study

Our main goal in this study is to identify fewer but more useful features for forecasting stack voltage compared to all the features in the original dataset. The original dataset comprises a total of 24 features. The question of whether the forecasting performance of these 24 features could be enhanced by utilizing a smaller feature set allowed us to conduct this study. To achieve this objective, feature selection is carried out using ChOA, and subsequently, stack voltage prediction is also performed using the selected features.

The ChOA algorithm reduces the number of features from 24 to 9, meaning that more than about 56% of the feature set size is reduced. Figure 3 displays the size of the feature set both before and after reduction.

ChOA algorithm selects U3, U4, U5, I, ToutH2, PinAIR, DinH2, DinAIR, and Utot features as more informative features. We perform a comparative analysis of the forecasting performance of BiLSTM and ChOABiLSTM models

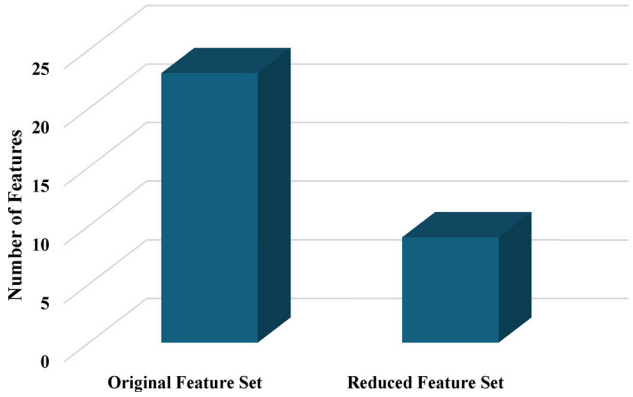


Fig. 3 Number of features before and after feature reduction

Table 3 Summary of models used for comparison

Model name	Optimizer	Activation function	#LSTM layer	# Hidden neurons	Epoch	Dropout	Dense	# Flatten
LSTM	Adam	Tanh	1	128	50	0.2	5, 1	1
ChOALSTM	Adam	Tanh	1	128	50	0.2	5, 1	1
StackedLSTM	Adam	Tanh	2	128, 128	50	0.2	5, 1	1
ChOAStackedLSTM	Adam	Tanh	2	128, 128	50	0.2	5, 1	1

Table 4 Performances of the BiLSTM and ChOABiLSTM models for test dataset

	Performance metrics		
Models	RMSE	MAE	R^2
BiLSTM	0.008	0.005	0.663
ChOABiLSTM	0.007	0.004	0.853

before and after feature selection for experimental dataset. The summary of models and comparative results are given in Tables 3 and 4, respectively.

From experimental viewpoint, ChOABiLSTM model reduces the error based on RMSE and MAE and achieves the better and acceptable R^2 compared to BiLSTM. Therefore, the reduced dataset demonstrate the effectiveness of the proposed feature selection method based on performance metrics and is superior to the original dataset.

When the loss curve of the training and test data is examined, a sharp decrease is observed in the curves. Then, the stability point, where no improvement in the fit of the model or decrease in loss is observed and reached in the eighth iteration as shown in Fig. 4.

Based on model parameters, expected and predicted values of Utot for BiLSTM and ChOABiLSTM models are represented with Fig. 5.

When comparing the expected and predicted values, the ChOABiLSTM model demonstrates a closer alignment with the expected outcomes than the BiLSTM model. Additionally, comparative experiments are conducted with ML algorithms, namely, support vector regression (SVR), ChOASVR, K nearest neighbor regression (KNNR) and ChOAKNNR, and DL algorithms, namely, LSTM, StackedLSTM, ChOALSTM, and ChOAStackedLSTM (LSTM + LSTM). The results are given in Table 5.

As it can be seen in Table 5, proposed ChOABiLSTM model performs better at all baselines and their chimp optimized version in terms of RMSE, MAE, and R^2 for experimental dataset. In our investigation, BiLSTM model provided dependable and satisfactory prediction outcomes. By providing both forward and backward time flow, BiLSTM can access future information in addition to the information at the current time point. This provides greater context and makes predictions more accurate. Moreover, it is evident that the high performance achieved using the reduced feature set selected by ChOA ensures that no significant information is lost, and key features are effectively identified.

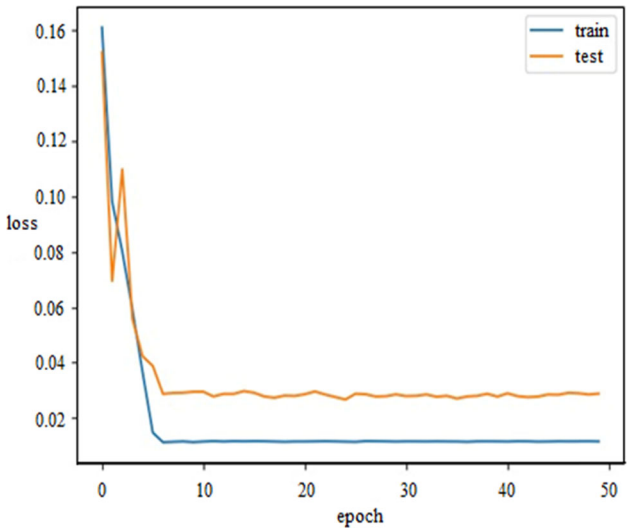
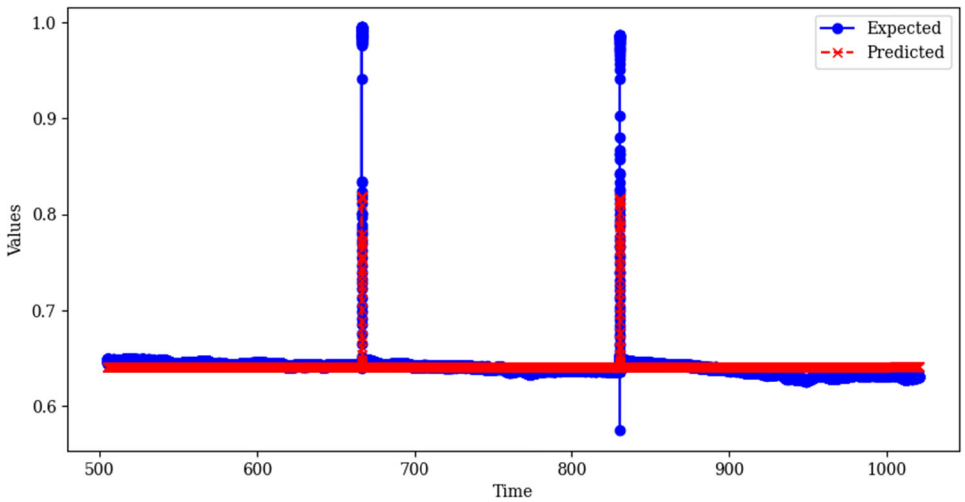
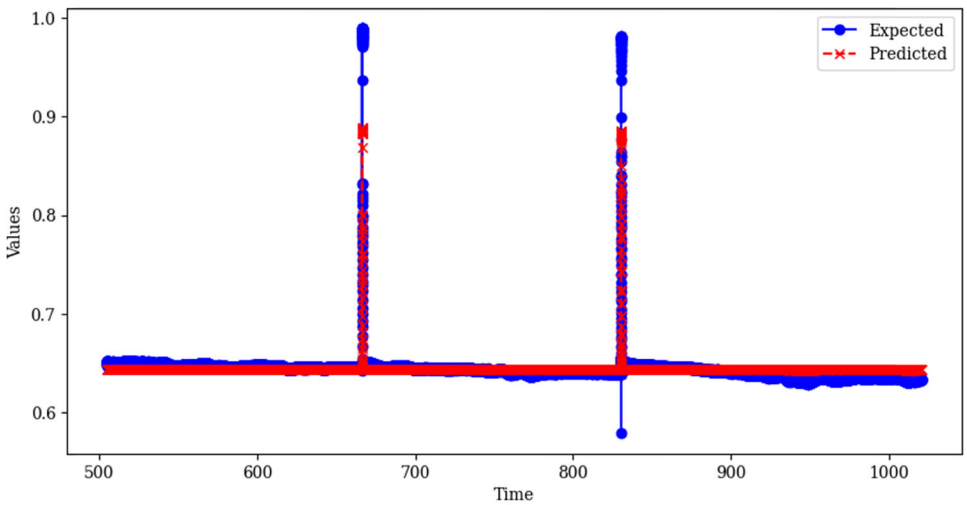


Fig. 4 Loss curve of the ChOABiLSTM model for train and test data



(a)



(b)

Fig. 5 Comparison between the expected and predicted Utot values for (**a**) BiLSTM and (**b**) ChOABiLSTM

Table 5 Performance of comparative algorithms on test set against ChOABiLSTM

Models	Performance metrics		
	RMSE	MAE	R^2
SVR	0.076	0.044	− 0.095
ChOASVR	0.064	0.045	0.222
KNNR	0.044	0.034	0.637
ChOAKNNR	0.031	0.023	0.821
LSTM	0.077	0.034	− 0.124
ChOALSTM	0.012	0.005	0.850
StackedLSTM	0.015	0.005	− 0.005
ChOAStackedLSTM	0.014	0.005	− 0.004
BiLSTM	0.008	0.005	0.663
ChOABiLSTM	0.007	0.004	0.853

3.4 Ablation study

Feature selection based on feature ranking has garnered significant attention from researchers. To evaluate the accuracy of our ChOABiLSTM model predictions, we employed metaheuristic algorithms such as simulated annealing (SA), cuckoo search (CS), and harmony search (HS), as well as feature ranking-based selection methods using XgBoost regression and random forest regression (RFR) algorithms, all combined with BiLSTM.

According to Table 6, it is seen that with ChOA-based selected features shows the best prediction performance. The ChOABiLSTM model, which was run with 9 features selected from 24 features, exhibited the best performance by reaching the lowest values (0.007 and 0.004) in terms of root-mean-square error (RMSE) and mean absolute error (MAE). This shows that the model provides high accuracy in its predictions. SABiLSTM, run with 15 feature selection, has relatively high error rates in terms of RMSE and MAE (0.095 and 0.066), and its performance is also supported by the negative correlation coefficient (− 0.672). This result shows that the model's predictions are less accurate. CSBiLSTM trained with 10 features is one of the model with the lowest performance and with high error rates (RMSE: 0.684 and MAE: 0.619) and a correlation value of − 1.210. It was observed that feature selection was not effective for this structure. The HSBiLSTM model obtained with 10 features produced relatively stable but low accuracy results with a moderate error rate (RMSE: 0.077, MAE: 0.032) and negative correlation value (− 0.100).

Table 6 Performance value comparison of the results with others for same dataset

Algorithms	Selected features		Performance metrics		
	Number	Features	RMSE	MAE	R^2
CHOABiLSTM	Total selected: 9/24	U3, U4, U5, I, ToutH2, PinAIR, DinH2, DinAIR, Utot	0.007	0.004	0.853
SABiLSTM	Total Selected: 15/24	U1, U2, U3, U4, I, TinAIR, ToutAIR, ToutWAT, PoutAIR, DoutH2, DinAIR, DoutAIR, DWAT, HrAIRFC, Utot	0.095	0.066	− 0.672
CSBiLSTM	Total Selected: 10/24	U3, U5, ToutH2, ToutAIR, TinWAT, DinH2, DoutH2, DoutAIR, DWAT, HrAIRFC	0.684	0.619	− 1.210
HSBiLSTM	Total selected: 10/24	U3, U5, I, TinH2, ToutAIR, ToutWAT, PoutH2, PinH2, DinH2, DinAIR	0.077	0.032	− 0.100
HHOBiLSTM	Total selected: 2/24	U1, DinAIR	0.075	0.030	− 0.061
XGBoostBiLSTM	Total selected: 9/24	ThinH2, Utot, PinAir, ToutWAT, PoutAIR, U1, TinAIR, DWAT, U2	0.015	0.005	− 0.002
RFRBiLSTM	Total selected: 8/24	U5, Utot, U4, U3, U2, U1, I, J	0.007	0.005	0.795

Table 7 Performance comparison of ChOABiLSTM and other models in the literature for experimental dataset

Models	RMSE	R^2
ChOALSTM	0.012	0.850
ChOAStackedLSTM	0.014	− 0.004
ChOABiLSTM	**0.007**	**0.853**
GA-BPNN [28]	–	0.419152
GA-TDNN [28]	–	0.571206
GA-NARX [28]	–	0.500815
W-NARX [4]	0.077	–
CEEMDeCNNeLSTM [30]	1.8005	–

Best values are bolded

Although the model run with only 2 feature selections showed good performance with low error rates (RMSE: 0.075 and MAE: 0.030), the correlation coefficient of − 0.061 limited the contribution of the small number of features to the accuracy. XGBoost supported BiLSTM model, which is run with 9 features, has low RMSE (0.015) and MAE (0.005) values. Correlation coefficient shows low negative correlation with − 0.002, indicating that features provide accurate predictions in this model, but correlation accuracy is limited. RFR supported BiLSTM exhibits low error rates (RMSE: 0.007, MAE: 0.005) and positive correlation value (0.795) indicating high accuracy with 8 feature selection. This result shows that RFR supported model provides strong accuracy with selected features. Also, as can be seen from Appendix A—Fig. S1, the most important feature is ThinH2. The remaining 8 important features are Utot, PinAir, ToutWAT, PoutAIR, U1, TinAIR, DWAT, and U2, respectively. Based on random forest regression, the most important features are U5 and Utot as seen in Appendix A—Fig. S2. U4, U3, U2, U1, I, and J are the remaining important features, respectively.

Compared with studies, the ChOABiLSTM model showed superior performance compared to models proposed in the other studies with both low error rate and high explanatory power (Table 7).

4 Conclusions

This research integrates the ChOA with LSTM models to predict performance degradation in PEM fuel cells. By analyzing the prediction results of CHOA variants and other methods, the following conclusions are drawn.

1. BiLSTM demonstrates efficacy in predicting voltage fluctuations in fuel cells. Compared with SVR, KNNR, LSTM, and StackedLSTM models with/without feature selection, the ChOABiLSTM model has a higher prediction accuracy, in which RMSE is 0.007, MAE is 0.004, and R^2 is 0.853 under the same conditions.
2. Effective feature selection significantly influences prediction outcomes, with the selection of suitable features yielding optimal performance degradation predictions. ChOA outperforms traditional feature selection algorithms like XGBoost and RFR by selecting the most informative features.
3. A high number of features in the vector will result in increased error rates and longer training times. Selecting informative features enhances accuracy.
4. The proposed ChOA-BiLSTM algorithm offers superior performance in PEMFC degradation prediction compared to other algorithms.
5. In the real-world operation of fuel cell vehicles, the fuel cell system operates continuously under varying loads, making the performance degradation process more complex. To resolve these issues, many researchers have concentrated on forecasting the performance of PEM fuel cells, enabling their integration into controllers to improve durability and efficiency in complex loading scenarios.
6. In further work, we plan to ensemble the proposed ChOABiLSTM model with other modalities to improve the forecasting accuracy of fuel cell performance degradation under complex operating conditions and extreme environment.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1140/epjs/s11734-024-01408-8.

Funding Not applicable.

Data availability The authors declare that the data supporting the findings of this study are publicly available.

Declarations

Conflict of interest The authors declare that no competing interests exist.

References

1. K. Benaggoune, 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**, 118835 (2022)
2. Caponetto R, Guarnera N, Matera F, Privitera E, Xibilia MG (2021) Application of electrochemical impedance spectroscopy for prediction of fuel cell degradation by LSTM neural networks. In: 2021 29th Mediterranean conference on control and automation (MED). IEEE, pp 1064–1069
3. D. Chen, W. Wu, K. Chang, Y. Li, P. Pei, X. Xu, Performance degradation prediction method of PEM fuel cells using bidirectional long short-term memory neural network based on Bayesian optimization. Energy **285**, 129469 (2023)
4. K. Chen, S. Laghrouche, A. Djerdir, Prognosis of fuel cell degradation under different applications using wavelet analysis and nonlinear autoregressive exogenous neural network. Renew. Energy **179**, 802–814 (2021)
5. K. Chen, S. Laghrouche, A. Djerdir, Degradation prediction of proton exchange membrane fuel cell based on grey neural network model and particle swarm optimization. Energy Convers. Manag. **195**, 810–818 (2019)
6. K. Chen, S. Laghrouche, A. Djerdir, Degradation model of proton exchange membrane fuel cell based on a novel hybrid method. Appl. Energy **252**, 113439 (2019)
7. Z. Deng, Q. Chen, L. Zhang, K. Zhou, Y. Zong, H. Liu, L. Ma, Degradation prediction of PEMFCs using stacked echo state network based on genetic algorithm optimization. IEEE Trans. Transp. Electrif. **8**(1), 1454–1466 (2021)
8. E. Ekinci, S.İ Omurca, B. Özbay, Comparative assessment of modeling deep learning networks for modeling ground-level ozone concentrations of pandemic lock-down period. Ecol. Model. **457**, 109676 (2021)
9. E. Ekinci, A comparative study of LSTM-ED architectures in forecasting day-ahead solar photovoltaic energy using weather data. Computing 1–22 (2024)
10. E. Ekinci, Ö. Ekinci, A. Morkoyunlu, Comparison of long short-term memory networks for daily ahead-flow predictions on a river. Environ. Eng. Manag. J. **23**(2), 287–300 (2024)
11. Z. Garip, E. Ekinci, A. Alan, Day-ahead solar photovoltaic energy forecasting based on weather data using LSTM networks: a comparative study for photovoltaic (PV) panels in Turkey. Electr. Eng. **105**(5), 3329–3345 (2023)
12. R. Ghasemlounia, A. Gharehbaghi, F. Ahmadi, H. Saadatnejadgharahassanlou, Developing a novel framework for forecasting groundwater level fluctuations using Bidirectional long short-term memory (BiLSTM) deep neural network. Comput. Electron. Agric. **191**, 106568 (2021)
13. A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. **18**(5–6), 602–610 (2005)
14. K. He, L. Mao, J. Yu, W. Huang, Q. He, L. Jackson, Long-term performance prediction of PEMFC based on LASSO-ESN. IEEE Trans. Instrum. Meas. **70**, 1–11 (2021)
15. K. He, Z. Liu, Y. Sun, L. Mao, S. Lu, Degradation prediction of proton exchange membrane fuel cell using auto-encoder based health indicator and long short-term memory network. Int. J. Hydrogen Energy **47**(82), 35055–35067 (2022)
16. E.H. Houssein, M.M. Emam, A.A. Ali, An efficient multilevel thresholding segmentation method for thermography breast cancer imaging based on improved chimp optimization algorithm. Expert Syst. Appl. **185**, 115651 (2021)
17. Z. Hua, Z. Zheng, M.C. Péra, F. Gao, Remaining useful life prediction of PEMFC systems based on the multi-input echo state network. Appl. Energy **265**, 114791 (2020)
18. I.K. Ihianle, A.O. Nwajana, S.H. Ebenuwa, R.I. Otuka, K. Owa, M.O. Orisatoki, A deep learning approach for human activities recognition from multimodal sensing devices. IEEE Access **8**, 179028–179038 (2020)
19. M.J. Izadi, P. Hassani, M. Raeesi, P. Ahmadi, A novel WaveNet-GRU deep learning model for PEM fuel cells degradation prediction based on transfer learning. Energy **293**, 130602 (2024)
20. C. Jia, H. He, J. Zhou, K. Li, J. Li, Z. Wei, A performance degradation prediction model for PEMFC based on bidirectional long short-term memory and multi-head self-attention mechanism. Int. J. Hydrogen Energy **60**, 133–146 (2024)
21. H. Jia, K. Sun, W. Zhang, X. Leng, An enhanced chimp optimization algorithm for continuous optimization domains. Complex Intell. Syst. 1–18 (2021)
22. M. Kaur, R. Kaur, N. Singh, G. Dhiman, Schoa: a newly fusion of sine and cosine with chimp optimization algorithm for hls of datapaths in digital filters and engineering applications. Eng. Comput. **38**(Suppl 2), 975–1003 (2022)
23. M. Khishe, M.R. Mosavi, Chimp optimization algorithm. Expert Syst. Appl. **149**, 113338 (2020)
24. 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. Hydrogen Energy **47**(78), 33466–33478 (2022)
25. J. Liu, Q. Li, W. Chen, Y. Yan, Y. Qiu, T. Cao, Remaining useful life prediction of PEMFC based on long short-term memory recurrent neural networks. Int. J. Hydrogen Energy **44**(11), 5470–5480 (2019)
26. L. Mao, L. Jackson, IEEE 2014 Data Challenge Data. Loughborough University. Dataset (2016)

27. R. Ranjbarzadeh, P. Zarbakhsh, A. Caputo, E.B. Tirkolaee, M. Bendechache, Brain tumor segmentation based on optimized convolutional neural network and improved chimp optimization algorithm. Comput. Biol. Med. **168**, 107723 (2024)
28. Y. Shen, M. Alzayed, H. Chaoui, Forecasting the remaining useful life of proton exchange membrane fuel cells by utilizing nonlinear autoregressive exogenous networks enhanced by genetic algorithms. J. Power Sources Adv. **24**, 100132 (2023)
29. T. Si, D.K. Patra, S. Mondal, P. Mukherjee, Breast DCE-MRI segmentation for lesion detection using chimp optimization algorithm. Expert Syst. Appl. **204**, 117481 (2022)
30. B. Sun, X. Liu, J. Wang, X. Wei, H. Yuan, H. Dai, Short-term performance degradation prediction of a commercial vehicle fuel cell system based on CNN and LSTM hybrid neural network. Int. J. Hydrogen Energy **48**(23), 8613–8628 (2023)
31. J.F. Torres, A. Galicia, A. Troncoso, F. Martínez-Álvarez, A scalable approach based on deep learning for big data time series forecasting. Integr. Comput. Aided Eng. **25**(4), 335–348 (2018)
32. F.K. Wang, X.B. Cheng, K.C. Hsiao, Stacked long short-term memory model for proton exchange membrane fuel cell systems degradation. J. Power. Sources **448**, 227591 (2020)
33. Y. Xie, J. Zou, Z. Li, F. Gao, C. Peng, A novel deep belief network and extreme learning machine based performance degradation prediction method for proton exchange membrane fuel cell. IEEE Access **8**, 176661–176675 (2020)
34. Y. Xie, J. Zou, C. Peng, Y. Zhu, F. Gao, A novel PEM fuel cell remaining useful life prediction method based on singular spectrum analysis and deep Gaussian processes. Int. J. Hydrogen Energy **45**(55), 30942–30956 (2020)
35. Y. Yang, X. Yu, W. Zhu, C. Xie, B. Zhao, L. Zhang, R. Zhang, Degradation prediction of proton exchange membrane fuel cells with model uncertainty quantification. Renew. Energy **219**, 119525 (2023)
36. D. Zhang, P. Baraldi, C. Cadet, N. Yousfi-Steiner, C. Bérenguer, E. Zio, An ensemble of models for integrating dependent sources of information for the prognosis of the remaining useful life of proton exchange membrane fuel cells. Mech. Syst. Signal Process. **124**, 479–501 (2019)
37. S. Zhang, T. Chen, F. Xiao, R. Zhang, Degradation prediction model of PEMFC based on multi-reservoir echo state network with mini reservoir. Int. J. Hydrogen Energy **47**(94), 40026–40040 (2022)
38. W. Zhu, B. Guo, Y. Li, Y. Yang, C. Xie, J. Jin, H.B. Gooi, Uncertainty quantification of proton-exchange-membrane fuel cells degradation prediction based on Bayesian-gated recurrent unit. eTransportation **16**, 100230 (2023)