Energy and AI 22 (2025) 100623

Contents lists available at ScienceDirect

Energy and AI

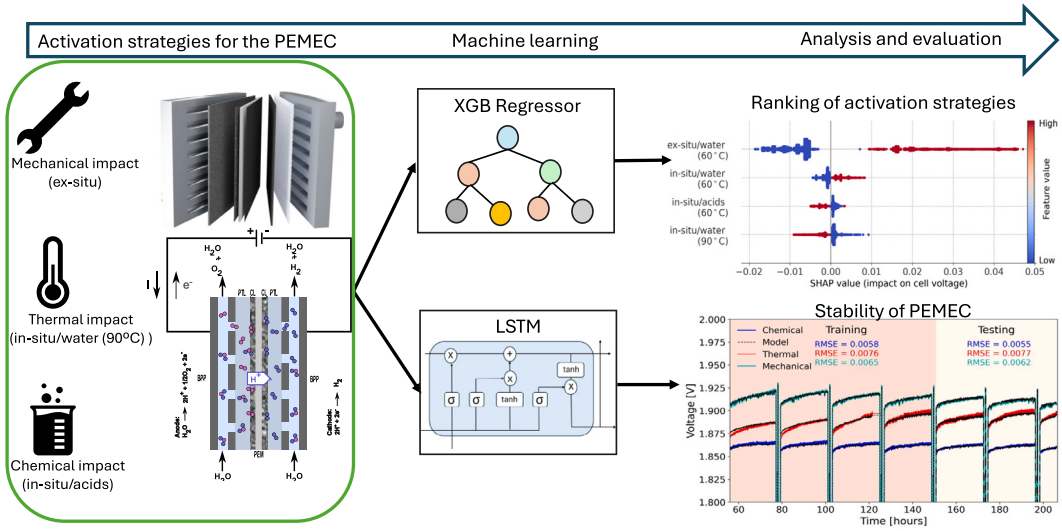
journal homepage: www.elsevier.com/locate/egyai

Evaluating activation strategies and their stability on PEM water electrolyzers using machine learning

K. Ashoke Raman [a,*], Niklas L. Wolf [a,b], Ali Javed [a], Violeta Karyofylli [a], Hans Kungl [a], André Karl [a], Eva Jodat [a], Rüdiger-A. Eichel [a,b,c]

[a] Institute of Energy Technologies, Fundamental Electrochemistry (IET-1), Forschungszentrum, Jülich, 52425, Germany
[b] Institute of Physical Chemistry, RWTH Aachen University, Aachen, 52062, Germany
[c] Faculty of Mechanical Engineering, RWTH Aachen University, Aachen, 52062, Germany

GRAPHICAL ABSTRACT



HIGHLIGHTS

- Introduced MEA activation as a feature in data-driven modeling of PEMEC.
- Combined data-driven models with SHAP for ranking MEA activation procedures.
- On the short run, *in-situ/water* 90 °C activation strategy showcases better performance.
- The LSTM model shows a good comparison with experimental data for cell stability .

ARTICLE INFO

Keywords:
Machine learning
Cell conditioning

ABSTRACT

Pre-treatment of the proton exchange membrane water electrolyzers is a crucial procedure performed prior to its regular operation. These procedures help in catalyst activation and membrane saturation, thereby, ensuring

* Corresponding author.
E-mail address: r.kuppa@fz-juelich.de (K.A. Raman).

https://doi.org/10.1016/j.egyai.2025.100623
Received 1 August 2025; Received in revised form 12 September 2025; Accepted 21 September 2025
Available online 4 October 2025

PEM water electrolyzer
Cell stability

its optimal performance. In this study, we use machine learning to investigate the impact of three distinct activation procedures on the cell performance and stability. The data set necessary to develop the surrogate models was obtained from a lab scale PEM electrolyzer cell. After evaluating the performance of the three tested models and validating them with experimental data, extreme gradient boosting is selected as the to perform parametric analysis. The modeling predictions reveal that the activation procedures mainly impact the ohmic resistance at the beginning of the cell life. These observations were further corroborated using through sensitivity analysis performed through an explainable artificial intelligence technique. Furthermore, data-driven time-series forecasting analysis to predict cell stability for different activation procedures showed a good comparison between experimental data and model predictions.

1. Introduction

Decarbonizing the energy landscape is vital not only for environmental preservation but also for long-term economic stability. This imperative has led to global efforts in the pursuit of green energy technologies [1]. In this endeavor, proton exchange membrane (PEM) water electrolysis has emerged as a leading technology for producing green hydrogen. PEM electrolyzers produce high purity hydrogen, operate at high current densities, and their modular design facilitates integration with existing transport and power grid infrastructure [2]. The central unit of these systems is the membrane electrode assembly (MEA), whose core functionalities include facilitating the electrochemical water splitting reactions, separation of gaseous products, and enabling proton and electron transport. As the critical electrochemical and transport processes occur in the MEA, it dictates the efficiency and long-term operational life of the cell. To improve the system's stability, reproducibility and performance before regular operation, the MEA is subjected to activation procedures during which current or thermal protocols are applied. These procedures [3] are also important in preventing accelerated degradation.

Before being deployed, PEMECs undergo activation procedures that enhance their performance. The duration of these activation procedures can range from less than an hour to over 20 h, even for identical materials. Understanding why these activation processes enhance the performance of PEM based electrochemical systems is both important and intriguing. During the manufacturing process of membranes, cations such as Fe^{3+}, Ni^{2+} and Cr^{2+} may contaminate the membrane and inhibit cell performance [4]. They do this by attacking the polymer bands of the membrane, resulting in mechanical and chemical destabilization. Some of these also react with protons to block and decrease protonic conductivity. The membrane is also exposed to different temperature and humidity conditions during its manufacturing process. Both these factors influence the water uptake of the membrane, thereby, causing varying swelling behaviors and changes in structural morphology, which in turn influence the proton conductivity, before regular cell operation [5,6]. Some of the catalyst particles are covered with oxide impurities or may contain solvents, used to prepare the catalyst paste, filling in the pores of the carbon paper. These factors effectively reduce the available active electrochemical surface area and result in poor cell performance. As such, activation procedures [7,8] help in optimizing cell start-up by activating the catalyst layer, improving membrane water saturation and through the removal of impurities.

Although conditioning procedures for PEMECs are well established [3], comparatively few comprehensive studies have investigated their effects relative to PEMFCs. A pre-conditioning procedure, in which the cell was subjected to a constant current density for 24 h was carried out to improve membrane hydration and stabilize the oxidation state of catalysts [9]. To standardize testing protocols and enable reliable comparison of PEM water electrolyzers across five laboratories, identical test cells, materials and operating protocols were used. For a given operating temperature and flow rate, a pre-conditioning protocol was carried out until the variation in current density was less than 1%. Across the different laboratories, a low maximum deviation in cell voltage was observed for different operating temperatures. Electrochemical characterization and in-situ visualization revealed that conditioning procedures increased the number of reaction sites [10]. Electrochemical impedance spectra (EIS) measurements revealed that both ohmic and overpotential losses decreased after conditioning. More recently, the impact of ex-situ and two different in-situ conditioning procedures was investigated to evaluate their effect on short-term cell performance [11]. While experiments provide critical information on the impact of conditioning strategies through electrochemical characterization techniques, there has been a recent shift towards integrating data-driven models with experimental data. To scale-up these systems for commercial utilization and to perform real-time monitoring, data-driven models have emerged as an essential tool for enabling predictive analytics, control, and optimization, thereby bridging the gap between laboratory-scale experimentation and industrial deployment [12–14].

In the domain of PEMECs, machine learning has been predominantly deployed to screen electrocatalysts [15–17]. Unlike PEMFCs, where substantial work [18] has been conducted on applying data-driven models for performance prediction and system-level optimization across various scales, similar efforts for PEMECs remain in their early stages of development [19,20]. Images obtained from X-ray based tomography of the anode PTL of PEMEC were used as training data to develop a machine model that predicted the oxygen content in the PTL [21]. The k-nearest neighbors (KNN) and decision tree regression models were used to optimize the design of PEMECs by focusing on selecting the best flow-field pattern which maximizes hydrogen production efficiency [22]. By using a combination of datasets generated from in-house experiments and literature, the authors demonstrated for the first time a data-driven framework for efficient hardware selection (flow field, catalyst type, membrane type, number of cells) for PEM water electrolyzers. Subsequently, design parameters were proposed for PEM electrolyzers with large scale hydrogen production rates ranging from 50–300 mL/min. A database consisting of data from 789 experiments extracted from 30 publications in the last 10 years was generated to find the most critical parameters influencing the performance of PEMECs [23]. By using a combination of ML based methods such as SHAP, KNN and Bayesian optimization, operating temperature, catalyst loading, ionomer content, support materials for anode and cathode, and appropriate pore structure of PTL were prescribed as the critical parameters. Similar data-driven cell design optimization was performed using polynomial regression [14]. A good agreement on the hydrogen production rate was observed between custom made cell using the parameters predicted by the data-driven model and the model's prediction. More recently, a data-driven methodology was proposed [24] to predict degradation trends of PEMECs by utilizing operational data. Ozdemir and Pektezel [25] investigated the effect of cell voltage, temperature, torque and flowrate on cell performance using machine learning algorithms from experimental data. They found that support vector machines outperformed the performance of random forest and multi-layer perceptron algorithms to predict current density and hydrogen flowrate. A low MAE of 0.006 and 0.0317 was noted for current density prediction on training and test data, respectively. The authors have also recently investigated PEMEC performance using synthetic data generated from numerical simulations [26]. Among the four different ML models they used, SVM demonstrated the highest

Nomenclature

ANN	Artificial Neural Network
CL	Catalyst Layer
CPE-T	Constant phase element - Exponent (P)
CPE-T	Constant phase element - Magnitude (T)
HER	Hydrogen Evolution Reaction
KNN	K Nearest Neighbors
LSTM	Long-Short Term Memory
MAE	Mean Absolute Error
MAE	Mean absolute error
MAPE	Mean Absolute Percentage Error
MEA	Membrane Electrode Assembly
ML	Machine Learning
MSE	Mean Square Error
OER	Oxygen Evolution Reaction
PEM	Proton Exchange Membrane
PEMEC	Proton Exchange Membrane Electrolytic Cell
PEMFC	Proton Exchange Membrane Fuel Cell
PTL	Porous Transport Layer
R^2	Coefficient of determination
RCT	Charge transfer resistance
RMSE	Root Mean Square Error
Rs	Solution resistance
SHAP	Shapely Additive Explanations
SR	Split Ratio
SVM	Support vector machines
XGB	Extreme Gradient Boosting

predictive performance with RMSE values of 0.0108 and 0.0371 on training and testing data, respectively.

From the above literature, it is evident that research in the direction of integrating PEMEC experiments and machine learning has been predominantly focused on optimizing material properties, cell design and operating conditions. A critical parameter, cell activation, which influences membrane hydration and catalyst activation, has not been considered yet as a predictive feature in these data-driven investigations. Additionally, it is imperative for economical feasibility to predict the stability of the electrolyzer when operating under these different conditioning protocols. This would assist in selecting a conditioning procedure that not only shows maximal short-term performance, but is also suitable for long term sustained efficiency. This study aims to fill these gaps of knowledge. Firstly, we build upon experimental data obtained from lab-scale PEMEC for three different conditioning strategies and develop data driven models that include conditioning as a feature. This allows the model to predict cell performance by considering the influence of MEA's initial treatment. Secondly, using SHAP as an explainable AI technique, we investigate which among the three pre-treatment procedures leads to the best cell performance. Finally, we model the stability of PEMEC using a database generated for over 200 h of cell operation. The temporal dependencies were captured using an LSTM. With its gated architecture, the LSTM circumvents the vanishing and exploding gradient problem encountered in conventional recurrent neural networks.

The paper is organized as follows: We begin with a brief description of the four different ML techniques 1(b) used for our work in Section 2. Data acquisition and information on data pre-processing are also provided in this section. Results are presented and discussed in Section 3 followed by conclusions in Section 4

2. Methodology

2.1. Data curation

The data were obtained from four in-house built single cells with commercial HYDRion MEAs purchased from Ion Power GmbH consisting of Nafion N115 membrane. The anode catalyst layer comprised iridium oxide with a loading of 1.0 mg/cm^2, while the cathode comprised platinum with a loading of 0.3 mg/cm^2. The active area of the MEA is 4.4×4.4 cm^2, which is larger than the active MEA area of 9 cm^2 used in a recent study [25]. A 250 μm-thick CURRENTO 2GDL10N-0.25 porous transport layer, with a platinum coating thickness of 0.25 μm on both sides, was used as the anode-side PTL. A Toray carbon paper (TGP-H-60) was used as the cathode PTL. The bipolar plates were fabricated from titanium grade 2 and featured a parallel flow-field design. The anode and cathode sides of these plates were coated with platinum (2.5 μm) and gold (2.5 μm), respectively. Temperature control was achieved by heating each side of the cell with a pair of heating rods, regulated by a single temperature sensor on each side. The cells were operated in test stands with high-purity water supplied at a flow rate of 0.05 L/min. After pre-treatment, the operating temperature was set to 60 °C. Further information on the experimental setup and corresponding test stations is provided in [11].

The input features consisted of applied current density, frequency and amplitude. A new feature indicating the type of pre-treatment conditioning was added to the list of features. *One-hot encoding* [27] was applied to the conditioning feature to convert it into numerical data. The output variables were the cell voltage, the real and imaginary components of impedance. Both features and output variables were scaled using min–max scaling. The input features and the measured output variables are listed in Table 1. The range along which the input parameters were varied and the output variables were measured are also provided in this table. Furthermore, the description of each activation strategy is shown in Fig. 1. For every experiment, the data were logged at an interval of one second. From each of the four datasets, the data corresponding to the pre-treatment procedure was removed. Four different current densities: 0.5, 1.0, 1.5, and 2.0 A/cm^2 were applied in ascending order for 12 h each. During these 12 h of operation at each current density, three polarization curves and impedance spectroscopy measurements were carried out. To prevent any outliers arising from spurious measurements, only data with voltage greater than or equal to 1.4 V were considered, as this corresponds to the open-circuit voltage. Screening of missing values was performed and corresponding cells were removed from the database. After cleaning and combining the data from the four cell experiments, the entire database constituted of 133 558 data points. The database exhibits high diversity with the categorical feature, *pre-treatment* being evenly distributed for all the four considered cases with the Shannon entropy measured at 1.997 bits out of 2. Similarly, with a standard deviation of 0.62 A/cm^2 and coefficient of variation of 1.03, the applied current density shows heterogeneity among the samples. In data-driven modeling, it is customary to split the data into training and test sets. Machine learning algorithms use the training data to learn the relationship between the input features and output variables. The test data are then used to evaluate the model's performance in predicting outcomes on this unseen data set. We have used a split ratio of 8:2 to randomly divide the experimental database into training and test sets. A small dataset comprising of one polarization curve and impedance measurements taken at 1.0 A/cm^2 was set aside to validate the trained models.

For predicting the electrolyzer stability, a separate database consisting of 200 h of operational data is recorded for each activation procedure. The temporal prediction for cell voltage employing the LSTM model is made using this database.

Before we proceed with building machine learning models, it is important to check if the type of conditioning strategy has any significant influence on the cell voltage. Statistical significance tests like analysis

Table 1
Range of measured variables provided to the data-driven model.

Parameters	Type	Unit	Range of values
Current density	Input	A/cm^2	[0.0–2.0]
Frequency	Input	Hz	[0.1–5 × 10^5]
Activation	Input (categorical)	–	ex-situ, in-situ/water(60 °C), in-situ/acids, in-situ/water(90 °C)
Voltage	Output	V	[1.4–1.93]
Impedance (Real)	Output	mΩ	[7.5–20.48]
Impedance (Imaginary)	Output	mΩ	[0–3.85]

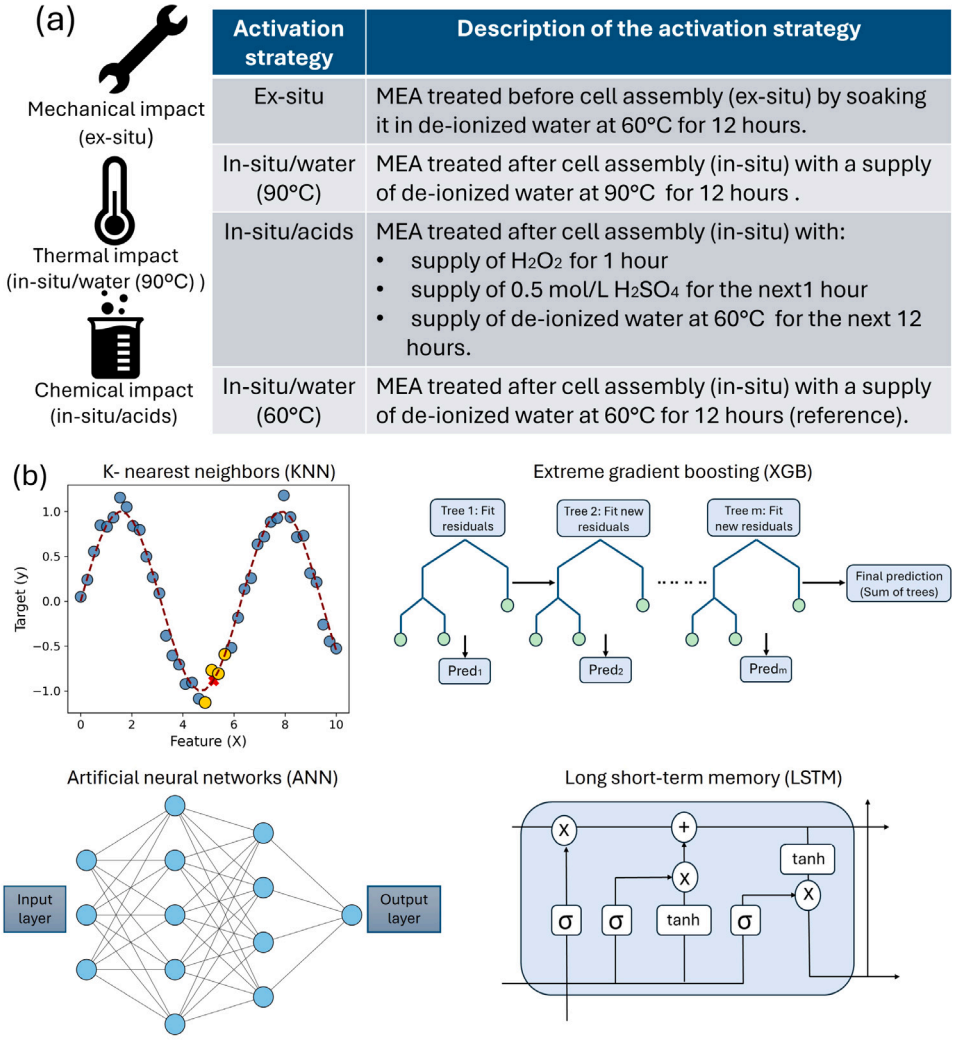


Fig. 1. (a) Description of the different MEA activation strategies investigated in this study. The *in-situ/water* (60 °C) is the reference activation strategy case against which the other pre-treatments methods were compared. (b) Schematic representation of the working principle of the employed machine learning models.

of variance (ANOVA) serve the important purpose of screening input features through evidence and help avoid feeding irrelevant variables or noise into the machine learning model. Since the focus of this study is to investigate the effect of conditioning on cell performance, we use ANOVA to verify its statistical significance. Table 2 shows the results of the one-way ANOVA performed on the cell voltage with the type of conditioning strategy as the categorical feature. With a p-value less than 0.001 and F(3, 133 554) = 11.81, the ANOVA reveals that the type of MEA conditioning has a significant effect on cell voltage. Therefore, it is important to consider it as a input feature for ML modeling. To confirm further, we performed a pariwise comparison of the different conditioning strategies using the Tukey HSD statistical significance test. Table 3 presents these comparisons with adjusted p-values and confidence intervals. The combinations reporting adjusted p-values less than 0.05 reject the null hypothesis. Based on the adjusted p-values and confidence intervals, we observe that *ex-situ* conditioning results in higher voltages compared to the *in-situ* methods. Moreover, the mean voltages for the base *in-situ* case are higher than the *in-situ/water* (90 °C) MEA conditioning. These results reiterate that the type of conditioning has an statistically significant effect on cell voltage.

2.2. K-Nearest Neighbors (KNN)

K-nearest neighbors [28] is a non-parametric machine learning algorithm which is employed to make regression predictions. It account for the values of near of nearby data points to make a prediction.

Table 2
One-way ANOVA results for the effect of conditioning on cell voltage.

Source	Sum of squares	df	F	p-value
Pre-treatment	0.703	3	11.81	<0.001
Residual	2649.82	133 554	–	–

Table 3
Tukey HSD pairwise comparisons of voltage (Ewe) for different pre-treatment types. Significant differences ($p < 0.05$) are indicated.

Group 1	Group 2	p-adj	Lower CI	Upper CI
ex-situ	in-situ/acids	0.007	−0.0063	−0.0007
ex-situ	in-situ/water(60°)	0.000	−0.0083	−0.0029
ex-situ	in-situ/water(90°)	0.965	−0.0033	0.0023
in-situ/acids	in-situ/water(60°)	0.223	−0.0049	0.0007
in-situ/acids	in-situ/water(90°)	0.042	0.0001	0.0059
in-situ/water(60°)	in-situ/water(90°)	0.000	0.0022	0.0079

Mathematically, the prediction made by KNN is expressed as:

$$\hat{y} = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i \tag{1}$$

where $\hat{y}$ is the prediction made by the model, k is the number of nearest neighbors, x_i are the data points which belong to the set N_k, x is the input where we want to predict the output and y_i is the target values corresponding to x_i. The hyperparameter to select for this model is the number of nearest neighbors (k). The algorithm proceeds with selecting the nearest k number of points, based on their euclidean distance with the input x. The final output prediction is made by taking the average of the target values at these k points from the training database.

2.3. Extreme gradient boosting (XGB)

XGBoost (eXtreme Gradient Boosting) is a machine learning technique which uses an ensemble of decision trees built in a sequential manner for solving both regression and classification problems. The working principle of a single decision tree involves recursively splitting the training dataset into subsets. This splitting is done based on selecting a particular input feature and its value such that it creates a homogeneous subset. This splitting process continues until the tree has reached a maximum depth or has a minimum number of samples in the resulting node. XGB operates by iteratively building a weak decision tree to predict the negative gradient of the loss function (squared error), which is the basis of the gradient boosting algorithm. Each of these decision trees is sequentially fit to minimize the residuals of the preceding trees.

2.4. Artificial neural network (ANN)

Inspired by the architecture and functionality of biological neurons, artificial neural networks are computational models which are capable of accurately predicting regression problems. The basic working unit of ANN is the neural node or neuron. Neurons are mathematical operators, which receive inputs and transmit output signals known as activations. A group of neurons are stacked together to form a layer. The first layer represents the input layer wherein, each node transmits individual input parameters. Similarly, the last layer is the output layer constituting the nodes which predict the target variables evaluated by the model. Interim layers are known as hidden layers which process and extract features from data. The information flow occurs from the input to the output layer through the hidden layers. Further details of XGB and ANN methods have been outlined in [19,20].

2.5. Long-short term memory (LSTM) model for forecasting

LSTM [29] are specialized version of recurrent neural networks (RNN) which overcome the problem of vanishing and exploding gradients, inherent to RNNs. In order to overcome these limitations, the LSTM networks have a gated architecture and memory cells which control the flow of information. These gates assist LSTM network in selectively retaining or discarding information. The LSTM architecture constitutes of the following components:

- **Cell state**: This component carries the information about the long term history of the time series. The cell state influences the output of the LSTM unit at the current time step.
- **Forget gate**: It controls the amount of the previous long-term memory that needs to be retained for the current time step. It employs a sigmoid function and is mathematically expressed as:

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{2}$$

- **Input gate**: This gate has two components comprising of a tanh layer and sigmoid layer. These two components determine the new long-term memory and the amount of this new long term memory that contributes to the information passed by the forget gate, respectively.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{3}$$

$$\hat{C}_t = \sigma(W_C.[h_{t-1}, x_t] + b_C) \tag{4}$$

$$C_t = (f_t \times C_{t-1}) + (i_t \times \hat{C}_t) \tag{5}$$

- **Output gate**: The last gate decides the output of the LSTM unit using the current cell state.

$$O_t = \sigma(W_o.[h_{t-1,x_t}] + b_o) \tag{6}$$

$$h_t = o_t \times tanh(C_t) \tag{7}$$

In the above mentioned equations, W_x and b_x correspond to the weights and biases of different gates ($x : f, i, C$ and o). A number of LSTM units could be stacked to form a layer, and such different layers can placed sequentially to allow the passage of information. To perform the stability analysis, we have used two layers having 50 and 25 LSTM units, respectively.

3. Results and discussion

We present the results by first assessing the performance of the three machine learning models using residual plots. We then follow up by validating these models using experimental data. We then perform a parametric analysis using the model which performs best on all the target variables.

3.1. Model performance

Residual plots allow us to visually infer the error distribution and patterns generated from the predictions made by the ML model. Fig. 2 illustrates the residual plots for KNN, XGB and ANN machine learning models. Each of these models was trained on three separate target variables, namely, cell voltage (top row), real part (middle row) and the negative imaginary part of impedance (bottom row). The search range and the values of the final selected hyperparameters for the trained ML models are provided in Table 4. Mathematically, the residual is defined as the difference between the experimental value and model prediction. For instance, the residual plots in the middle row show the accuracy of the three models in predicting the real component of impedance when compared against the actual experimental data. While cell voltage provides insights into the overall efficiency of the cell, impedance measurements shed light in segregating the different overlapping processes, such as electrode kinetics, ohmic resistance and mass transfer losses occurring inside the system. The residuals obtained

Table 4
Hyperparameters and search range for each machine learning model.

Model	Hyperparameter	Search range	Selected values (Voltage, Re(Z) and -Im(Z))
KNN	n_neighbors	[3,8,11,14]	3, 8 and 8
XGB	n_estimators	[100–800]	300, 800 and 800
	learning_rate	[0.05–0.20]	0.015, 0.005 and 0.005
ANN	No. of hidden layers	[2–4]	4, 3 and 4
	No. of hidden units	[20–100]	[100-80-60-30], [60-40-20] and [100-80-60-30]
	learning rate	[0.0001–0.01]	0.00025, 0.001 and 0.00015

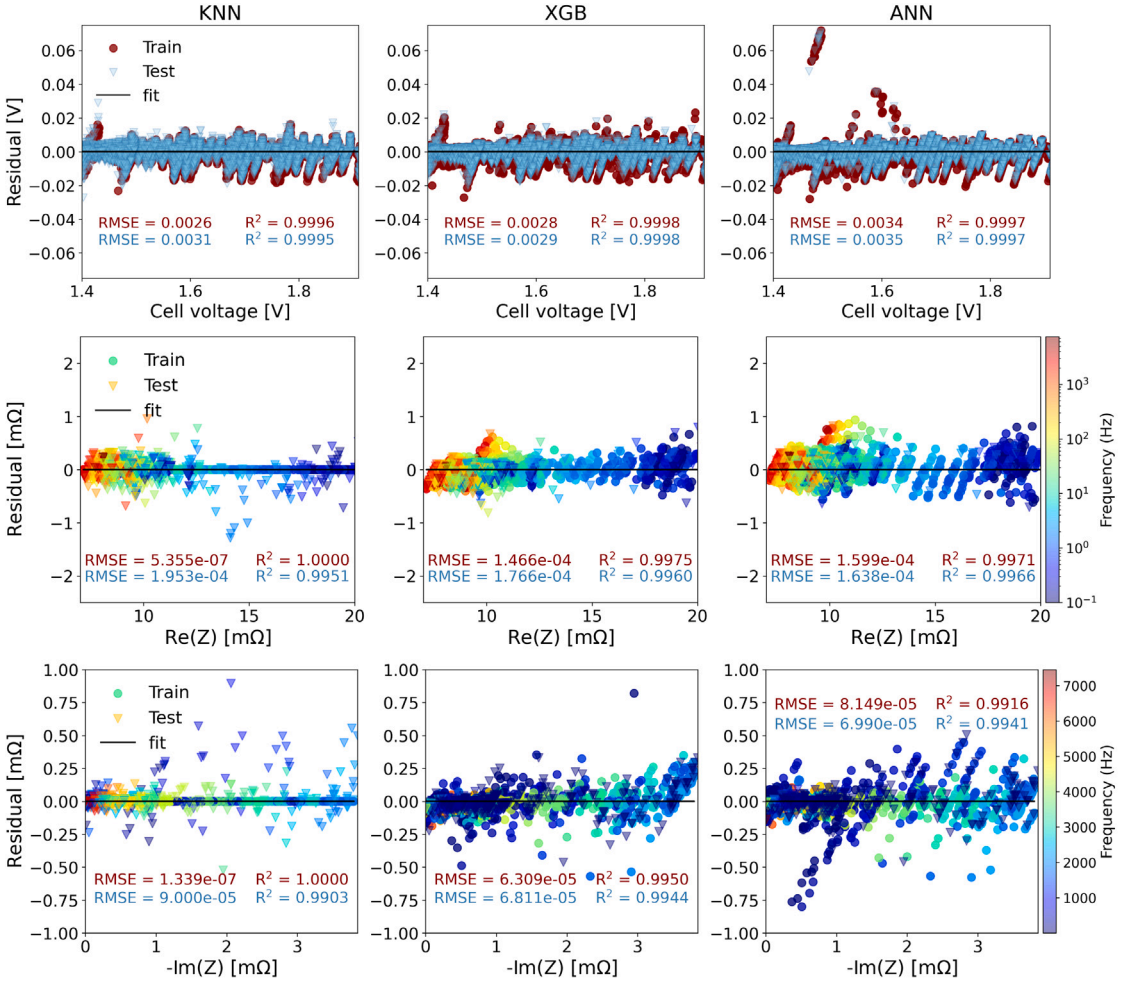


Fig. 2. Residual plots showing the comparison between the predicted current densities (top row), real (middle row) and imaginary (bottom row) components of impedance with those obtained from experimental data. The model's performance on the training and test set are shown in each plot by the RMSE and R^2 values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

from the training and testing datasets are indicated using different colors. Furthermore, the root mean square error (RMSE) and coefficient of determination (R^2) are used to quantify the performance of the machine learning models. These metrics are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2} \tag{8}$$

$$R^2 = 1 - \frac{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}{\frac{1}{n}\sum_{i=1}^{n}(y_j - \bar{y}_j)^2} \tag{9}$$

where n denotes the number of samples, y_j represents the ground truth obtained from the experimental data, $\hat{y}$ denotes the model prediction and $\bar{y}$ is the average value of y_j.

From Fig. 2 (top row), we observe that the residual distribution across the entire range of cell voltage is similar for all the three tested models. The residuals for both training and test data are within 0.02 V for KNN and XGB models. Some outliers, especially for the test data, are observed for the ANN model, for which the residuals reach as large as 0.06 V. The KNN model shows the best performance on the training data with RMSE = 0.0026, whereas for the unseen test data, XGB shows a slightly better performance with RMSE = 0.0031. All the models have R^2 greater than 0.999, indicating good model performance in predicting voltage.

While voltage prediction is directly linked to applied current density, electrochemical impedance spectroscopy (EIS) is a complex valued variable that is frequency dependent. Several system parameters such as membrane resistance, double-layer capacitance and diffusion, which represent various processes transpiring within the system, influence impedance. These factors impose challenges in modeling impedance accurately. We infer this behavior from the distribution of residuals

Table 5
Comparison of methods with performance metrics on Voltage, real (Re(Z)), and imaginary (-Im(Z)) components of impedance on the test data.

Method	Split type	Voltage		Re(Z)		-Im(Z)	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
KNN	Random	0.999	3.0×10^{-3}	0.994	2.11×10^{-4}	0.983	1.15×10^{-4}
	Grouped	0.975	2.3×10^{-2}	0.802	1.3×10^{-3}	0.953	1.91×10^{-4}
XGB	Random	0.982	1.84×10^{-2}	0.995	1.76×10^{-4}	0.994	6.81×10^{-5}
	Grouped	0.974	2.43×10^{-2}	0.722	1.54×10^{-3}	0.959	1.79×10^{-4}
ANN	Random	0.999	3.52×10^{-3}	0.996	1.63×10^{-4}	0.994	6.99×10^{-5}
	Grouped	0.963	2.87×10^{-2}	0.743	1.48×10^{-3}	0.714	4.74×10^{-4}

of the real (middle row) and imaginary (bottom row) components of impedance shown in Fig. 2. The KNN model shows an ideal fit for both Re(Z) and Im(Z) on the training data, with RMSE = 5.3×10^{-7} and 1.33×10^{-7} for the training and test data, respectively. However, the model's performance on the test data is marked by a pattern which is characterized by an increase in the spread of residuals as the applied frequency increases. This pattern is also observed for XGB and ANN models, with the residuals of Re(Z) for the ANN model showing a wavy pattern on the test data. The non-constant variance of residuals for impedance predictions indicates *heteroskedasticity* in the models. While the KNN model exhibits high variance due to overfitting the test data, the ANN model is characterized by a large number of outliers on the data for Im(Z). Although the RMSE values on the test data for both components of impedance indicate that ANN model performs slightly better than XGB, ANN model also showcases a higher degree of *heteroskedasticity* compared to XGB. For Re(Z), XGB model shows a lower variance in residuals when compared to the ANN model. In general, we observe that modeling the capacitive part (imaginary) of impedance is challenging when compared to the resistive component. Therefore, among the three tested machine learning models, the XGB model shows overall better performance for voltage and impedance prediction.

The model performance shown so far is based on random splitting of the database into training and testing data sets. Random splitting ensures that there is good representation of all feature values, both numerical and categorical features, in both datasets. This provides insights into the overall fitting ability of the models. However, as we are using the type of MEA activation as an input feature, the chances of *data leakage* may arise if the same specific activation type is present in both the training and test datasets. *Data leakage* occurs when measurements from the same experiment appear in both the training and test datasets. To investigate the occurrence of *data leakage* further, Table 5 shows the R^2 and RMSE performance metrics for all the three tested models under random and grouped splitting, with the test data under consideration. The group splitting was performed such that all samples from a specific activation type are kept only in the test dataset, while the training dataset consists of samples from the remaining activation type. We observe that R^2 and RMSE for voltage predictions show a slight drop when using grouped splitting compared to random splitting. This is a general observation, as the models are making predictions on unseen categorical features, which reflects their generalization ability. A sharp drop in the R^2 values (0.2–0.3) are noted for RE(Z) prediction when modeled under grouped splitting. KNN performs slightly better and XGB and ANN models. For Im(Z) predictions, ANN model exhibits overfitting as a large drop from 0.994 to 0.714 is observed when the splitting type is changed from random to group. The metrics shown in Table 5 indicate that the voltage predictions made by the models are relatively stable with respect to *data leakage* when compared to predicting impedance.

3.2. Model validation

The residuals and performance metrics give us insights into the overall performance of the trained models. To further elucidate the model performance quantitatively, we compare their predictions with experimental data for (a) thermal and (b) chemical conditioning procedures as shown in Fig. 3. Polarization curves and Nyquist plots measured at a current density of 1 A/cm^2 are used to draw the comparison. For the thermal conditioning procedure, the experimentally obtained polarization curve is in good agreement with all the three models as illustrated in Fig. 3(a). The Nyquist plot on the other hand exemplifies the differences in model predictions from experimental measurements. The fitting values obtained from the equivalent circuit model are shown in Table 6. The ANN model shows larger deviations with experimental data when compared to XGB and KNN models. For the thermal pre-treatment, the ANN model under predicts the ohmic resistance and continues this trend across the entire range of applied frequencies, marked by a shift in the Nyquist plot towards the left side of the experimental data. Predictions made by the KNN model show good agreement with impedance measurements in the high-to-medium frequency range. However, we notice that at low frequencies, it exhibits large deviations from experimental data (Re(Z) 9.0). Both the ohmic and charge transfer resistances are captured accurately by the XGB model. We observe that for the entire range of applied frequencies, the predictions made by the XGB model are close to the experimental data, albeit showing a slight overprediction in the capacitive component of impedance at medium frequencies. Fig. 3(b) elucidates the comparison between ML model predictions and experimental measurements for the case of applying chemical conditioning to the MEA. Similar to our earlier observation for thermal conditioning, the ML models show good agreement with experimental data for the polarization curve. Predictions made by both KNN and XGB models agree well with experimental measurements of impedance for chemical conditioning. However, in this case the ANN model shows slight overprediction of the ohmic and charge transfer resistances as it shifts slightly to the right of the experimental Nyquist plot.

From the residual distributions on the impedance data shown in Fig. 2, we conclude that the KNN model fits well to the training data and under performs when subjected to test data. Even though the model shows good agreement with experimental measurements in Fig. 3, we observe that the model tends to memorize the training data due to its instance based learning approach. The predictions of the model are highly dependent on the local structure of the data. This is shown as the model predicts noisy impedance data at low frequency for chemical conditioning, as noticed in Fig. 3(b). The lack of regularization techniques makes KNN prone to over-fitting. Neural networks on the other hand are designed with a lot of parameters which enable them to learn the complex underlying relationship between the input features and the components of impedance measurements. In order to establish this relationship accurately, ANNs require a large amount of training data. Especially for impedance measurements, where different electrochemical processes take place across a wide range of frequencies, the model can overfit or only capture the dominant mechanism when trained on a smaller data set. These factors may attribute to the under and over prediction of the ANN model on the experimental impedance measurements shown in Fig. 3(a) and (b), respectively. In terms of overall accuracy and effectively capturing the trends in experimental data, XGB has demonstrated better performance. Therefore, we select this model to proceed ahead with parametric analysis.

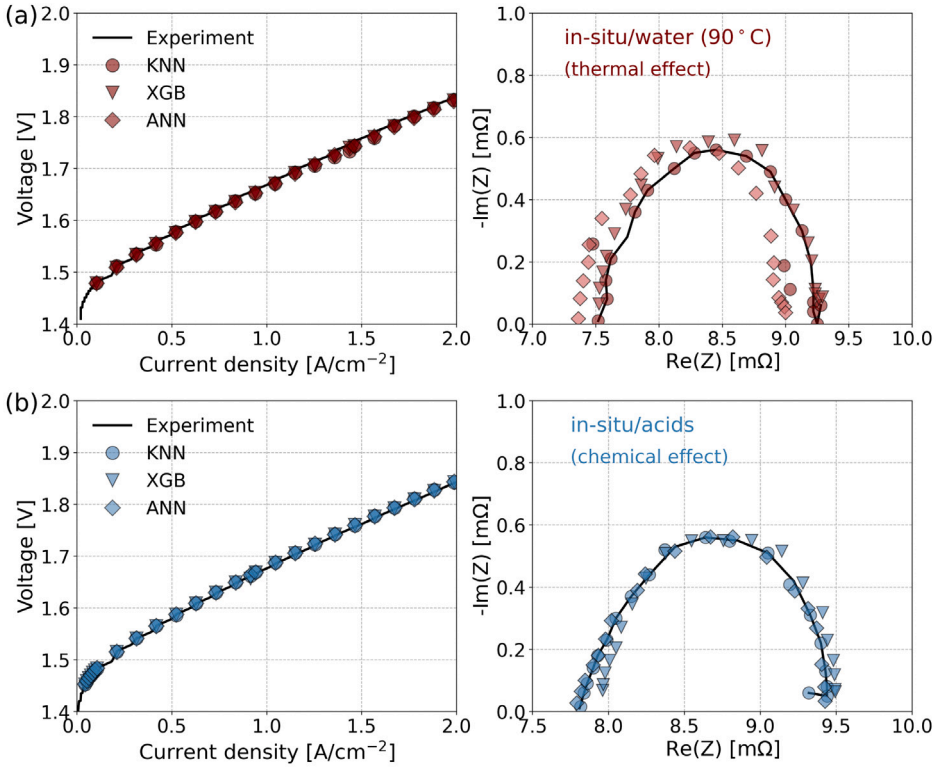


Fig. 3. Comparison between experimental measurements and model prediction through polarization curve and Nyquist plot for MEAs pre-treated with (a) deionized water at 90 °C and (b) acidic solutions. The impedance measurements were made at a current density of 1.0 A/cm^2. Additionally, it is to be noted that as the applied frequency increases, the real component of impedance decreases.

Table 6
Equivalent circuit model fitting parameters for different ML models corresponding to the Nyquist plot in Fig. 3.

Pre-treatment	Model	Rs (Ω)		RCT (Ω)		CPE-T		CPE-P	
		Value	Error%	Value	Error%	Value	Error%	Value	Error%
Thermal	Experiment	7.528	0.20	1.755	1.45	0.0076	7.74	0.7439	1.92
	KNN	7.500	0.38	1.719	2.67	0.0067	14.72	0.7599	3.53
	XGB	7.463	0.16	1.855	1.14	0.0095	5.78	0.7182	1.52
	ANN	7.340	0.21	1.676	1.55	0.0074	8.43	0.7624	2.04
Chemical	Experiment	7.832	0.12	1.682	1.19	0.0088	6.19	0.7542	1.51
	KNN	7.834	0.15	1.679	1.40	0.0091	7.25	0.7538	1.78
	XGB	7.921	0.15	1.666	1.46	0.0092	7.56	0.7389	1.87
	ANN	7.813	0.16	1.695	1.47	0.0093	7.57	0.7409	1.88

3.3. Effect of activation strategies

We now compare the influence of *ex-situ*, *in-situ/acids* and *in-situ/water (90 °C)* activation procedures against the *reference* case using electrochemical characterization. Polarization curves and Nyquist plots are used to elucidate the comparison among these protocols. All the results presented in this section are obtained from the XGB model. In this study, four different MEA activation procedures were carried out to evaluate their influence on PEMEC performance. One of these four pre-treatment strategies is selected as the *reference* case, against which other cases are compared. For the *reference* conditioning procedure, the MEA was hydrated with deionized water at 60 °C for 12 h. This procedure was performed after the MEA had been assembled in the cell. We now discuss the impact of other three activation procedures on cell performance and compare them with the *reference* case.

3.3.1. Mechanical impact

For the *ex-situ* activation procedure, the MEA was first hydrated with deionized water at 60 °C for 12 h (ex-situ) and then assembled into the cell in a wet state. This highlights the effect of membrane swelling, due to hydration, on the cell performance. Fig. 4(a) illustrates the comparison between *ex-situ* and *reference* conditioning procedures through polarization curves. For lower current densities (0–0.5 A/cm^2), we do not notice any substantial differences between the pre-treatment procedures. As the current density increases, we observe higher cell voltage values for *ex-situ* activation when compared to *reference* scenario. It can be seen from the inset figure that this difference in cell voltage values increases, especially at higher current densities. For example, a difference of 28.5 mV is observed at an applied current density of 2.0 A/cm^2 between these two conditioning procedures. The influence of different current densities on the impedance spectra for mechanical impact is shown through the Nyquist plot in Fig. 4(b). It is important to note that certain selected current densities (0.75, 1.25 and 1.75 A/cm^2) were not present in the original training data. This is intended to showcase the model's capability in making predictions for parameter values not included in the training database. Therefore, we use the developed data-driven model to make these predictions. The ohmic resistance is obtained from the high frequency limit of the impedance spectra. We notice a negligible change of 0.34 mΩ in ohmic resistance as current density varies from 2.0 to 0.50 A/cm^2. The

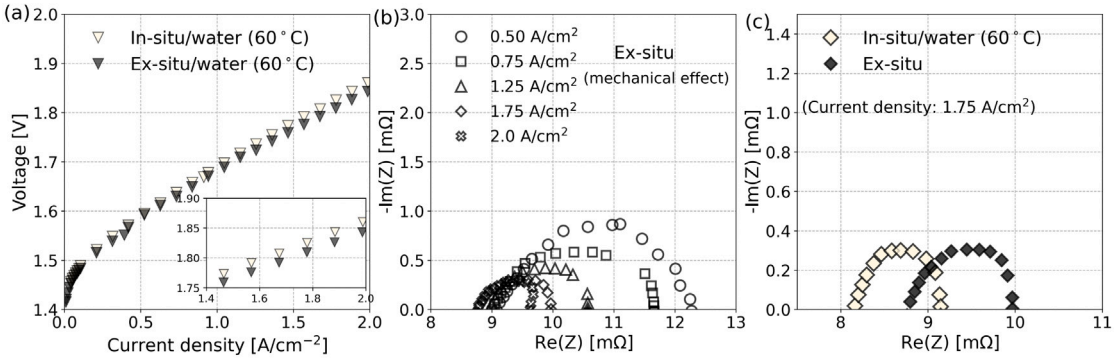


Fig. 4. Electrochemical performance comparison between *reference* and *ex-situ* activation. (a) Polarization curves highlighting performance reduction due to mechanical impact. (b) Nyquist plots at different current densities (0.5–2.0 A/cm^2) for *ex-situ* activation showcasing the changes in ohmic (R_Ω) and charge transfer resistance. (c) Comparison of Nyquist plots between *reference* and *ex-situ* activation, illustrating increased R_Ω due to mechanical impact.

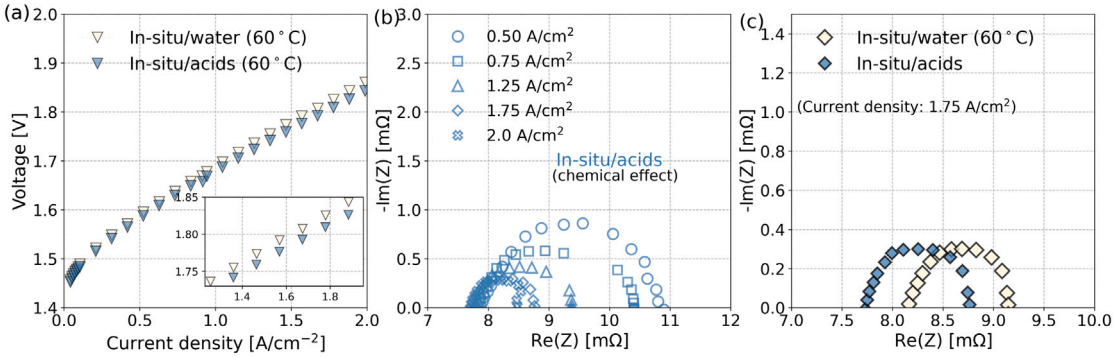


Fig. 5. Electrochemical performance comparison between *reference* and *in-situ/acids* activation. (a) Polarization curves highlighting performance improvement due to chemical impact. (b) Nyquist plots at different current densities (0.5–2.0 A/cm^2) for *in-situ/acids* activation showcasing the changes in ohmic (R_Ω) and charge transfer resistance. (c) Comparison of Nyquist plots between *reference* and *in-situ/acids* conditioning, illustrating reduced R_Ω due to chemical impact.

charge transfer resistance, obtained from the low frequency intercept, decreases as current density increases. This is attributed to the fact that as current density increases, the over potential increases, leading to an increase in the reaction driving force. As a result, the charge transfer resistance decreases. To exemplify the influence of mechanical impact on cell performance, we compare the Nyquist plots obtained at current density of 1.75 A/cm^2 in Fig. 4(c). The ohmic resistance for *ex-situ* and *reference* were found to be 8.79 mΩ and 8.16 mΩ, respectively. Similarly, the charge transfer resistance is increased from 9.14 mΩ to 9.96 mΩ, when the MEA underwent mechanical impact. These results indicate that *ex-situ* activation results in reduced cell performance.

3.3.2. Chemical impact

For *in-situ/acids* activation, the MEA was first assembled into the cell (*in-situ*) and treated with a continuous supply of hydrogen peroxide (H_2O_2) for 1 h. This procedure was then followed up by a consecutive supply of deionized water and 0.5 mol/L sulphuric acid (H_2SO_4) for an interval of 1 h each. Finally, the MEA was treated with deionized water at 60 °C for 12 h to conclude the *in-situ/acids* activation procedure. Fig. 5(a) illustrates the effect of *in-situ/acids* activation against *reference* on the polarization curve. Unlike the previous case shown in Fig. 4(a), we observe a slight improvement (in the order of millivolts) in voltage losses when the MEA undergoes pre-treatment chemically. This improvement is further exemplified from the inset shown in Fig. 5(a). Fig. 5(b) shows the effect of current density on the ohmic and charge transfer resistance for chemically treated MEA. A negligible change of 0.1 mΩ in ohmic resistance is observed when the current density increases from 2.0–0.5 A/cm^2. This indicates a lower variance in ohmic resistance for *in-situ/acids* activation when compared to *ex-situ* pre-treatment discussed above. However, the evolution of charge transfer resistance follows a similar trend, decreasing as current density increases. A 27.4% increase in charge transfer resistance is observed when the current density varies from 0.5–2.0 A/cm^2. Chemical pre-treatment of the MEA leads to reduction in the ohmic resistance, thereby, facilitating better mobility of ions across the membrane. This is inferred from the reduction of R_Ω from 8.16 mΩ for *in-situ/acids* activation to 7.75 mΩ for the *reference* case, as shown in Fig. 5(c). Along the same lines, the charge transfer resistance also reduces from 9.14 mΩ to 8.76 mΩ when the MEA is subjected to chemical pre-treatment.

3.3.3. Thermal impact

For the last conditioning procedure, the MEA was assembled in the cell (*in-situ*) and deionized water at 90 °C was supplied through the cell for 12 h. The difference between this procedure and the *reference* conditioning procedure lies in the temperature of the deionized water used during pre-treatment. During cell operation, measurements for all the conditioning procedures were made with water at 60 °C. Similar to the chemical activation procedure, MEA pre-treated with 90 °C shows improved performance, evidenced by a decrease in cell voltage, as illustrated from the polarization curve in Fig. 6(a). This reduction in cell voltage is observed across the range of applied current densities. Furthermore, the magnitude of this voltage reduction remains nearly constant across different current densities. For instance, the voltage differences between the *reference* and *in-situ/water(90 °C)* activation procedures at current densities 0.52 A/cm^2, 1.54 A/cm^2 and 1.95 A/cm^2 is found to be 14.61 mV, 14.8 mV and 13.78 mV, respectively. From Fig. 6(b), we observe that R_Ω remains largely unchanged as the applied current density increases from 0.5–2.0 A/cm^2. This maybe attributed to the fact that after pre-treatment with deionized water at 90 °C the membrane is fully hydrated, thereby, minimizing the

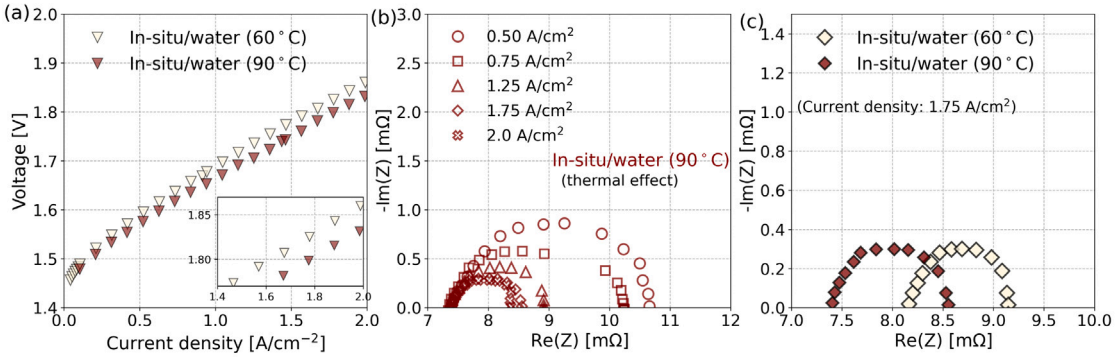


Fig. 6. Electrochemical performance comparison between *reference* and *in-situ/water(90 °C)* activation. (a) Polarization curves highlighting performance improvement for thermally activated MEA. (b) Nyquist plots at different current densities (0.5–2.0 A/cm^2) for *in-situ/water(90 °C)* activation showcasing the changes in ohmic (R_Ω) and charge transfer resistance. (c) Comparison of Nyquist plots between *reference* and *in-situ/water(90 °C)* activation, illustrating reduced R_Ω for thermal activation.

resistance for movement of H^+ through the membrane. As the charge transfer resistance decreases, the current density scales logarithmically with the overpotential. The impact of *in-situ/water(90 °C)* activation in assisting ion transfer through the membrane can be further corroborated from Fig. 6(c), where we observe that a decrease in R_Ω from 8.16 mΩ (*reference*) to 7.14 mΩ for an applied current density of 1.75 A/cm^2.

3.4. Comparative analysis of conditioning strategies using SHAP

In the preceding section, we compared the electrochemical performance comparison of the three conditioning strategies with that of the *reference* case, in which the MEA was hydrated with deionized water at 60 °C for 12 h. While the differences between each of these conditioning procedures were quantitatively illustrated against the *reference* case, an overall comparison between these procedures is yet to be provided. Moreover, the underlying variations in cell performance resulting from different conditioning treatments warrant further conjecture and discussion. We undertake these analysis in this section using the explainable AI technique known as SHAP.

The SHAP [30] technique is a method based on game theory that asserts importance values for each input feature that contributes to a given prediction made by the machine learning model. The SHAP value indicates the average contribution of a particular feature, across all possible input parameter combinations, in making model predictions. This in turn helps us to understand how and why the machine learning model favors a particular feature over others. Fig. 7(a) shows the SHAP plot elucidating which among the three conditioning procedures contributes most to the cell voltage. The x-axis represents the scale for SHAP values, and each point corresponds to an experimental instance, indicating the SHAP value for the respectively conditioning experiment. The ranking of conditioning strategies is based on their average SHAP values. The scale bar placed on the left of the plot represents the value of the features. It should be noted the unlike input parameters with numerical values, such as current density or frequency, categorical features do not have numerical values. These categorical features are encoded into numerical format using *one-hot encoding*. It transforms categories into binary vectors wherein each category is represented by a separate column. After performing *one-hot-encoding*, the single feature column of pre-treatment type consisting of four cases is transformed into four columns leading to the formation of three new features. Each of these four column corresponds to a specific pre-treatment type and has a binary value of 0 or 1. For instance, when a data point corresponding to the *ex-situ* is considered, the *ex-situ* feature column has the feature value = 1 and other three columns have the feature value = 0. In this way, *one-hot-encoding* transforms categorical features into binary numerical features. Therefore, each of the four categories (including *reference*) appear separately in Fig. 7(a). A red dot (binary value = 1) denotes the activation or presence of the corresponding conditioning process for a given instance, whereas a blue dot (binary value = 0) signifies its absence. We observe that the *ex-situ* procedure has the highest impact on cell voltage. When the conditioning process is activated, it predominantly has positive SHAP values ranging as far as 0.05. This indicates that *ex-situ* activation has a positive correlation with cell voltage, which means, it attributes to an unfavorable increase in cell voltage. This pre-treatment is followed by the base or the *reference* conditioning procedure. Similar to *ex-situ* activation, it also shows positive correlation with cell voltage. However, its impact on cell voltage is lower than that exerted by mechanical impact, as denoted by its SHAP values. A positive correlation emphasizes the fact that employing *ex-situ* activation leads to an increase in cell voltage, indicating higher voltage losses against the base case. A reversal in this trend is observed for these two other conditioning procedures. The SHAP value distribution for the *in-situ/acids* activation denotes that it has a negative correlation with cell voltage and the average SHAP value lower than the base case. Therefore, inline with our observations from the preceding sections, SHAP analysis also corroborate that chemically activated MEAs results in overcoming cell voltage losses. The *in-situ/water(90 °C)* activation procedure shows relatively better performance among the tested activation procedures.

From the above discussions, it is evident that *ex-situ* activation performs inferior to the base case, whereas the *in-situ/acids* and *in-situ/water(90 °C)* activation exhibit improved performance against the *reference*. These observed variations in cell performance could stem from various underlying factors. For *ex-situ* pre-treatment, the membrane is hydrated for 12 h *ex-situ* before it is assembled into the cell. While for *ex-situ* and *reference* cases the membrane is subject with water at same temperature, the topology of the membrane after hydration will differ in these two situations. For *reference* condition [11,31], the membrane is hydrated with the cell which is confined within the confinements of the cell geometry. Therefore, the predominant direction of membrane swelling would occur along the axial or through-plane direction. This enhances the contact between the catalyst particles and the porous PTL layers. However, owing to unrestricted membrane expansion in the case of *ex-situ* activation, the membrane undergoes swelling in both along the axial and through-plane directions. This expansion along the in-plane direction may hinder the contact between catalyst particles and PTL, once the membrane is assembled into the cell. This leads to an increase in R_Ω resulting in lower performance when compared to the base case. Pre-treatment with H_2O_2 is known to remove organic impurities from the catalyst layer [32] of MEAs manufactured in industry. This results in improving the contact resistance. Additionally, pre-treatment with H_2SO_4 enhances protonic conductivity as H^+ ions from the acid replace metal cations present in the membrane.

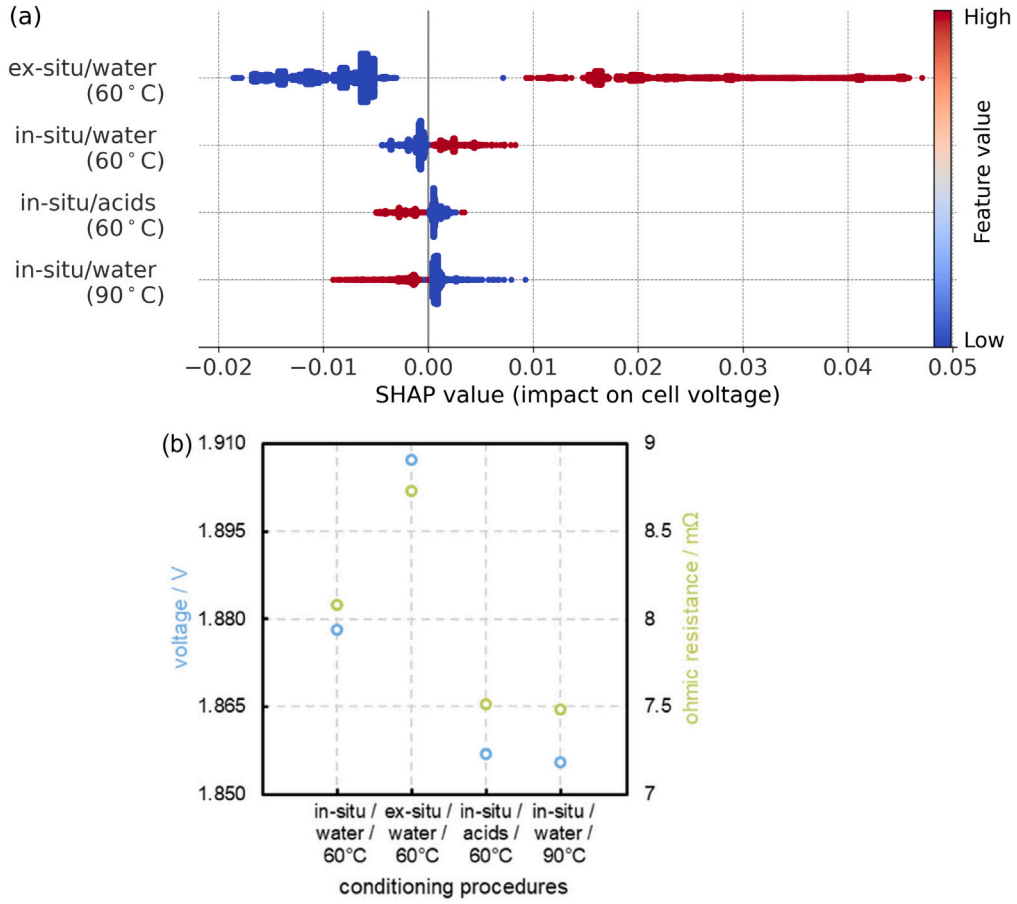


Fig. 7. (a) Ranking different MEA conditioning procedures using an explainable methodology (SHAP). The ranking is estimated using the developed XGB data-driven model. (b) Impact of conditioning procedure on cell voltage and ohmic resistance at 2 A/cm^2 using direct experimental data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Source: Figure adapted with permission from [11].

As the number of H^+ ions increases, the protonic conductivity increases. Therefore, the combined influence of lower interfacial resistance and enhanced protonic conductivity(in the ionomer/membrane) can be attributed to the improved performance of *in-situ/acids* activation against the base case. Finally, it is well established that when Nafion membrane is hydrated with deionized water at higher temperature, it increases the hydration state of the membrane [33]. This implies that the number of water molecules surrounding the sulphonic sites of the membrane increases. As a result, the ionic conductivity of the membrane increases during *in-situ/water(90 °C)* activation procedure leading to improved cell performance.

At this point, it is pertinent to note that SHAP analysis allows us to get insights into how the machine learning model makes its decision to rank the features based on their importance. However, it is important to verify that the validity of this interpretation is in line with actual experimental data. This in turn, would prevent us from drawing false conclusions and imbibe more trust on the model. Fig. 7(b) shows the comparison between all the conditioning procedures obtained from experimental measurements [11]. The voltage distribution at a current density of 2 A/cm^2 clearly indicates that *ex-situ* conditioning process leads to higher cell voltage when compared to *in-situ* at 60 °C. The *in-situ* acids shows better performance and *in-situ* water at 90° show better performance than the base case. The strong alignment between experimental rankings and SHAP analysis strengthens our trust in the reliability of SHAP interpretations.

From the above discussions, we have seen the impact of three activation procedures against a reference case. Each of these methods has its own advantages and disadvantages. While soaking a membrane *ex-situ* [11] enhances water uptake and initial membrane ionic conductivity, excessive swelling hinders catalyst loading accuracy [34]. Our results shown in Fig. 4 confirm this negative effect of mechanical impact on cell performance. Pre-treatment with chemicals removes organic/inorganic contaminant and manufacturing impurities. This leads to improved cell performance [11]. However, further exploration in terms of evaluating the effectiveness is still warranted by using spectroscopic techniques. Similarly, improved cell performance through thermal pre-treatment is well established in PEMFCs [8]? and more recently in PEMEC [11]. Our model predictions discussed in the previous section are inline with these observations. Nevertheless, the long term impact of thermal pre-treatment needs to be further investigated.

3.5. Effect of pre-treatment on electrolyzer stability using LSTM

Probing into the stability of cell voltage for a PEMEC, operating at a given current density, can reveal insights into the long-term performance of the system. It can provide insights into degradation effects, membrane hydration and catalyst utilization. Consistency in voltage output over a long-period of time indicates healthy cell operation. We now discuss the voltage stability analysis and compare the three pre-treatment procedures. Fig. 8 compares the voltage stability curves for *ex-situ*, *in-situ/water(90 °C)* and *in-situ/acids* activation procedures. The cell voltage is recorded at an operating current density of 2 A/cm^2 for a time interval from 58 to 206 h. Prior to reaching its current operating current density, the cell was operated at a series of lower

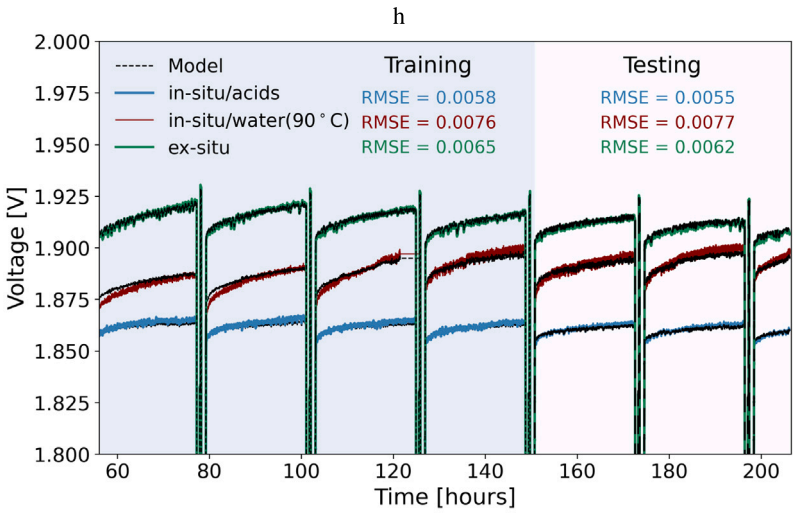


Fig. 8. Cell voltage profiles for around 150 h of PEMEC operation at an applied current density of 2 A/cm^2. The colored lines and the black dashed lines represent the experimental data for the three pre-treatment procedures and model predictions, respectively. The background color grading indicates the data used for training and testing intervals. The performance of the LSTM model is noted through the RSME values. The comparison reveals distinct stability characteristics across the conditioning strategies. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

current densities, which were gradually increased over time. Moreover, nearly after every 20 h, EIS and polarization curve measurements were conducted to monitor the performance at varying current densities. All the three conditioning methods showed stable voltage profiles thought the 150 h on operation. For instance, the cell voltage for *ex-situ* activation varied from 1.923 V to 1.921 V between the initiation of the first and second break-in procedure at 77.5 and 101.3 h, respectively. Before the commencement of the final break-in procedure at 196.7 h, a cell voltage of 1.913 was recorded. At the end of this operational time period which includes the intermittent procedures, a cell voltage drop of less than 10 mV was noticed. The other two conditioning procedures also exhibited similar pattern in voltage stability, showcasing the operational durability. A voltage difference of 5 mV and 17 mV were noted for *in-situ/acids* and *in-situ/water(90 °C)* activation cases between the beginning of the first and the last break-in procedures. In terms of the comparative performances for between the conditioning procedures, we observe that inline with our previous observations, the *ex-situ* pre-treatment showcases highest increase in cell voltage. *In-situ/acids* activation procedure showed more stable cell performance when considered for long-term operation. While the difference in the pre-break-in voltage between *ex-situ* and *in-situ/water(90 °C)* treatments tends to decrease at certain time intervals, specifically at 77.5 h and 196 h, the chemically treated MEA maintains a nearly constant cell performance, with voltage of 1.865 V and 1.863 V at these respective time instants. The predictions made by the LSTM model for the three cases are also shown in Fig. 8. The model shows good agreement on both the training and testing time intervals with the experimental data, indicating its capability to accurately capture the underlying dynamics. Since the change in cell voltage for chemically activated MEA is nearly constant throughout the time interval, the LSTM model shows lowest RMSE for this case with the values of 0.0058 and 0.0055 for training and test data, respectively. the RMSE values on the testing dataset for *ex-situ* and *in-situ/water(90 °C)* pre-treatment procedures were computed to be 0.0062 and 0.0077, respectively.

At this point, it is important to note that the effect of activation strategies can be different on the short-term and long term cell performance. In the previous section, we have observed that the activation induced by thermal effect performs slightly better than that applied with chemical impact. However, we not note a switch in this trend from our observations in Fig. 8. We clearly observe that for long-term operation, activation induced by chemical impact has a better performance than *in-situ/water (90 °C)* activation. While activating MEA with water at 90 °C increases its hydration state initially, as the cell continues to operate with water at 60 °C, its hydration state gradually comes down. Therefore, the cell voltage increases with time and its showcases a reduced performance when compared against the MEA activated chemically. To further elucidate this switch in performance behavior, Figure 10 illustrates the time period during which we notice this switch in performance between these two activation processes. Before 50 h, thermal impact shows a better performance than chemical impact. However, we notice change in this pattern with the time frame of 50–54 h. Beyond this time frame, chemically activated MEA show better performance.

A time series data constitutes of key components which can be additive or multiplicative: *trend*, *seasonality*, *cyclicity* and *residual*. These components inform us about the different features of the time series data. The *trend* component indicates the general direction of the time series data observed for a long period of time. The regular and periodic fluctuations in the data are characterized by its *seasonal* component. When these fluctuations in data are not periodic, they are related to the *cyclic* component. Finally, the random and unpredictable variations in the time series are governed by the *residuals*. Conventional time series methods [35] focus on predicting the *trend* and the *seasonal* component. The ease in extracting these components could be attributed to the fact that *trend* and *seasonal/cyclic* are characterized by a systematic and repeating patterns in data progression. Therefore, conventional statistical tools were designed to extract these patterns and repeating structures, while assuming the residuals as unpredictable white noise. However, it is crucial to model at least parts of the residual, to make reliable forecasting in real time applications.

Fig. 9 shows the comparison of the (a) *trend*(b) *seasonal* and (c) *residual* components of experimental data and cell voltage predictions made by the LSTM model for the three conditioning procedures. The time series were decomposed using additive assumption. In general, we observe the voltage has an increasing trend for all the three cases, albeit at a lower rate. This pattern in trend evolution repeats each time after the break-in procedure. A good agreement between the experimental and model prediction is noted. The more irregular but periodic *seasonal* component shown in Fig. 9(b) is also well captured by the LSTM model. In particular, we observe that the *seasonal* component for *in-situ/water(90 °C)* activation has a larger contribution to the overall cell voltage, with an amplitude variation of ±0.25 V when compared with the other two cases with ±0.05 V. The central advantage of the LSTM model is not only in its ability to accurately capture the *trend* and

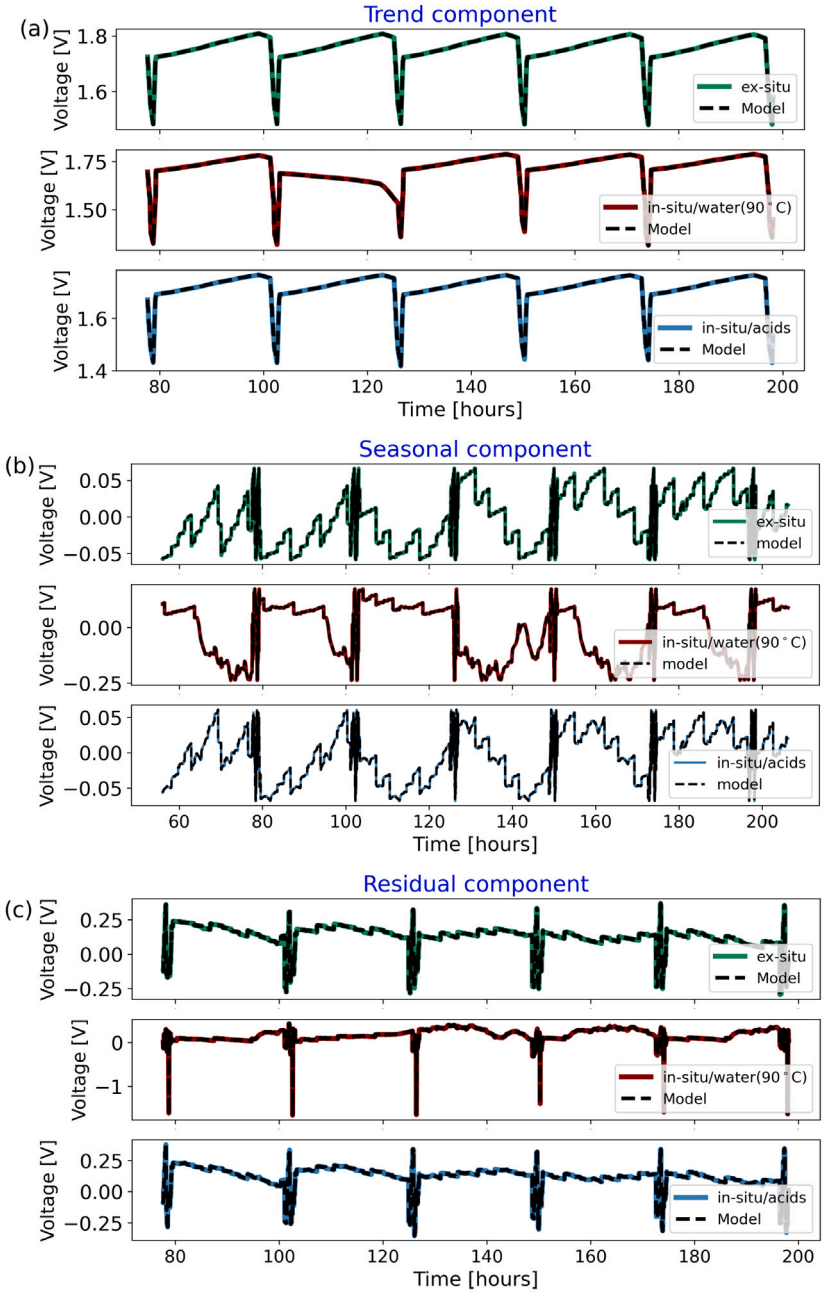


Fig. 9. Comparison of LSTM model predictions with experimental data across (a) trend (b) seasonal and (c) residual components of the cell voltage time series. The time series decomposition for the three conditioning procedures is shown, highlighting the model's ability to accurately capture the long-term trend, periodic seasonal patterns, and residual variations.

seasonal component, but also the *residuals* accurately as illustrated in Fig. 9(c). For all the three cases, we observe that the *residual* component has a larger contribution than the *seasonal* component to the voltage time series. The ability of LSTM to predict *residuals* lies with the assumption made in defining *residuals* by conventional methods [35]. While *residuals* are considered as purely random and uncorrelated, they often contain interactions with other variables, undetected auto-correlations and nonlinear dependencies. Deep learning models, such as LSTM, are capable of extracting and learning these non-linear patterns using memory cells and gates. They remember patterns from the past, even a random spike in the data, to make predictions. Moreover, influence of external factors, such as the applied current density in this case, is taken by the LSTM model to predict the residual part. Therefore, these data-driven models are capable to make improvised time-based voltage predictions.

4. Conclusions

The effect of four pre-treatment procedures on the performance of PEMEC was investigated through three different machine-learning methods: KNN, XGB, and ANN. A PEMEC comprising of Nafion N115 membrane with an active area of 4.4×4.4 cm^2 was used to generate the data required for model development. Since the focus of this work is to investigate the effect of conditioning procedures on cell performance,

the operating conditions and other parameters were held constant. The model performance was calibrated using the RMSE and R^2 performance metrics. The main conclusions drawn are as follows:

(1) All three models showed good performance in predicting the cell voltage for varying current densities and pre-treatment procedures. However, deviations in model performance for impedance predictions were noted. The KNN model showcased variance by performing well on the training data and over-fitting on the test data. Lack of sufficient EIS data hinders the performance of ANN model for impedance predictions. The XGB model showed an overall better performance for both cell voltage and impedance.
(2) *Ex-situ* conditioning leads to membrane swelling along both through plane and in-plane directions. Therefore, it showcases lower performance and higher ohmic resistance when compared to the *reference* scenario. Both *in-situ/acids* and *in-situ/water(90°)* activation procedures showed lower ohmic resistances and cell voltage when compared against the base case.
(3) A strong alignment between machine learning predictions and experimental results is observed, as SHAP analysis identified *in-situ/water(90°)* activation as the most effective, followed by *in-situ/acids*, *reference* and *ex-situ* for short-run conditioning operation. This ranking is consistent with experimental findings [11].
(4) Time series predictions from the LSTM model showed good agreement with experimental data. Unlike conventional forecasting techniques, LSTM is capable of making accurate predictions of the *residual* component of the time series.

Selecting suitable conditioning protocol is essential not only for obtaining optimal cell performance, but also to prolong electrolyzer health. The effect of certain pre-treatment condition is also depends on the duration of time-period considered. For short-term observations, we noticed that thermally activated MEAs perform slightly better than those activated chemically. However, from the stability analysis it is evident that *in-situ/acids* activation outperforms *in-situ/water(90°)* activation. In this context, data-driven models play a crucial role in identifying the optimal conditioning procedures, thereby minimizing both time and experimental costs.

CRediT authorship contribution statement

K. Ashoke Raman: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Niklas L. Wolf:** Writing – review & editing, Data curation. **Ali Javed:** Writing – review & editing, Formal analysis. **Violeta Karyofylli:** Writing – review & editing, Investigation, Formal analysis. **Hans Kungl:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **André Karl:** Writing – review & editing, Project administration. **Eva Jodat:** Writing – review & editing, Project administration. **Rüdiger-A. Eichel:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

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

Acknowledgments

The authors gratefully acknowledge the financial support by the German Federal Ministry of Education and Research (BMBF) within the H_2Giga project DERIEL (grant number 03HY122C). The authors gratefully acknowledge computing time on the supercomputer JURECA at the Forschungszentrum Jülich under the project no. PAJ2403. The authors would also like to thank Sebastian B. C. Lehmann for his contribution to the graphical designs and Zheng Jiang for data processing which was used in performing equivalent circuit modeling.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.egyai.2025.100623.

Data availability

Data will be made available on request.

References

[1] Rasul MG, Hazrat MA, Sattar MA, Jahirul MI, Shearer MJ. The future of hydrogen: Challenges on production, storage and applications. Energy Convers Manage 2022;272:116326.
[2] Wang Yun, Pang Yiheng, Xu Hui, Martinez Andrew, Chen Ken S. PEM fuel cell and electrolysis cell technologies and hydrogen infrastructure development–a review. Energy Environ Sci 2022;15(6):2288–328.
[3] Christmann Kläre, Friedrich K Andreas, Zamel Nada. Activation mechanisms in the catalyst coated membrane of PEM fuel cells. Prog Energy Combust Sci 2021;85:100924.
[4] Zamel Nada, Li Xianguo. Effect of contaminants on polymer electrolyte membrane fuel cells. Prog Energy Combust Sci 2011;37(3):292–329.
[5] Kusoglu Ahmet, Dursch Thomas J, Weber Adam Z. Nanostructure/swelling relationships of bulk and thin-film PFSA ionomers. Adv Funct Mater 2016;26(27):4961–75.
[6] Xie Tao, Page Kirt A, Eastman Scott A. Strain-based temperature memory effect for nafion and its molecular origins. Adv Funct Mater 2011;21(11):2057–66.
[7] Zhiani Mohammad, Majidi Somayeh, Silva Valter Bruno, Gharibi Hussein. Comparison of the performance and EIS (electrochemical impedance spectroscopy) response of an activated PEMFC (proton exchange membrane fuel cell) under low and high thermal and pressure stresses. Energy 2016;97:560–7.
[8] Qi Zhigang, Kaufman Arthur. Activation of low temperature PEM fuel cells. J Power Sources 2002;111(1):181–4.
[9] Siracusano Stefania, Trocino Stefano, Briguglio Nicola, Pantò Fabiola, Aricò Antonino S. Analysis of performance degradation during steady-state and load-thermal cycles of proton exchange membrane water electrolysis cells. J Power Sources 2020;468:228390.
[10] Wang Weitian, Li Kui, Ding Lei, Yu Shule, Xie Zhiqiang, Cullen David A, Yu Haoran, Bender Guido, Kang Zhenye, Wrubel Jacob A, et al. Exploring the impacts of conditioning on proton exchange membrane electrolyzers by in situ visualization and electrochemistry characterization. ACS Appl Mater Interfaces 2022;14(7):9002–12.
[11] Wolf Niklas, Javed Ali, Treutlein Leander, Kungl Hans, Karl André, Jodat Eva, Eichel Rüdiger-A. Tuning proton exchange membrane electrolytic cell performance by conditioning nafion N115-based membrane electrode assemblies. Electrochem Sci Adv 2025;e202400038.
[12] Chen Xia, Rex Alexander, Woelke Janis, Eckert Christoph, Bensmann Boris, Hanke-Rauschenbach Richard, Geyer Philipp. Machine learning in proton exchange membrane water electrolysis—A knowledge-integrated framework. Appl Energy 2024;371:123550.
[13] Ding Rui, Chen Yawen, Rui Zhiyan, Hua Kang, Wu Yongkang, Li Xiaoke, Duan Xiao, Li Jia, Wang Xuebin, Liu Jianguo. Machine learning utilized for the development of proton exchange membrane electrolyzers. J Power Sources 2023;556:232389.
[14] Mohamed Amira, Ibrahem Hatem, Yang Rui, Kim Kibum. Optimization of proton exchange membrane electrolyzer cell design using machine learning. Energies 2022;15(18):6657.
[15] Chen Letian, Tian Yun, Hu Xu, Yao Sai, Lu Zhengyu, Chen Suya, Zhang Xu, Zhou Zhen. A universal machine learning framework for electrocatalyst innovation: a case study of discovering alloys for hydrogen evolution reaction. Adv Funct Mater 2022;32(47):2208418.
[16] Timmermann Jakob, Kraushofer Florian, Resch Nikolaus, Li Peigang, Wang Yu, Mao Zhiqiang, Riva Michele, Lee Yonghyuk, Staacke Carsten, Schmid Michael, et al. IrO 2 surface complexions identified through machine learning and surface investigations. Phys Rev Lett 2020;125(20):206101.
[17] Flores Raul A, Paolucci Christopher, Winther Kirsten T, Jain Ankit, Torres Jose Antonio Garrido, Aykol Muratahan, Montoya Joseph, Nørskov Jens K, Bajdich Michal, Bligaard Thomas. Active learning accelerated discovery of stable iridium oxide polymorphs for the oxygen evolution reaction. Chem Mater 2020;32(13):5854–63.
[18] Ding Rui, Zhang Shiqiao, Chen Yawen, Rui Zhiyan, Hua Kang, Wu Yongkang, Li Xiaoke, Duan Xiao, Wang Xuebin, Li Jia, et al. Application of machine learning in optimizing proton exchange membrane fuel cells: a review. Energy AI 2022;9:100170.
[19] Raman K Ashoke, Hammacher Linus, Kungl Hans, Karl André, Jodat Eva, Eichel Rüdiger-A, Karyofylli Violeta. Data-driven surrogate modeling for performance prediction and sensitivity analysis of transport properties in proton exchange membrane water electrolyzers. Appl Energy 2025;386:125529.

[20] Karyofylli Violeta, Raman K Ashoke, Hammacher Linus, Danner Yannik, Kungl Hans, Karl André, Jodat Eva, Eichel Rüdiger-A. Elucidating parasitic currents in proton-exchange-membrane electrolytic cells via physics-based and data-driven modeling. Electrochem Sci Adv 2025;e70000.
[21] Satjaritanun Pongsarun, O'Brien Maeve, Kulkarni Devashish, Shimpalee Sirivatch, Capuano Cristopher, Ayers Katherine E, Danilovic Nemanja, Parkinson Dilworth Y, Zenyuk Iryna V. Observation of preferential pathways for oxygen removal through porous transport layers of polymer electrolyte water electrolyzers. Iscience 2020;23(12).
[22] Yang Rui, Mohamed Amira, Kim Kibum. Optimal design and flow-field pattern selection of proton exchange membrane electrolyzers using artificial intelligence. Energy 2023;264:126135.
[23] Günay M Erdem, Tapan N Alper, Akkoç Gizem. Analysis and modeling of high-performance polymer electrolyte membrane electrolyzers by machine learning. Int J Hydrog Energy 2022;47(4):2134–51.
[24] Woelke Janis, Rex Alexander, Eckert Christoph, Bensmann Boris, Hanke-Rauschenbach Richard. Predicting future polarization curves from operating data: Machine learning-based investigation of degradation modeling concepts for PEM water electrolysis. Energy AI 2025;100547.
[25] Ozdemir Safiye Nur, Pektezel Oguzhan. Performance prediction of experimental PEM electrolyzer using machine learning algorithms. Fuel 2024;378:132853.
[26] Ozdemir Safiye Nur, Pektezel Oguzhan. A CFD-driven machine learning approach for predicting PEM electrolyzer performance. Fuel 2026;405:136479.
[27] Harris Charles R, Millman K Jarrod, Van Der Walt Stéfan J, Gommers Ralf, Virtanen Pauli, Cournapeau David, Wieser Eric, Taylor Julian, Berg Sebastian, Smith Nathaniel J, et al. Array programming with NumPy. Nature 2020;585(7825):357–62.
[28] Hastie Trevor, Tibshirani Robert, Friedman Jerome H, Friedman Jerome H. The elements of statistical learning: data mining, inference, and prediction. vol. 2, Springer; 2009.
[29] Hochreiter Sepp, Schmidhuber Jürgen. Long short-term memory. Neural Comput 1997;9(8):1735–80.
[30] Lundberg Scott M, Erion Gabriel, Chen Hugh, DeGrave Alex, Prutkin Jordan M, Nair Bala, Katz Ronit, Himmelfarb Jonathan, Bansal Nisha, Lee Su-In. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020;2(1):56–67.
[31] Treutlein Leander, Javed Ali, Wolf Niklas L, Kungl Hans, Karl André, Jodat Eva, Eichel Rüdiger-A. Gas crossover in membrane electrolyzers - The impact of MEA conditioning on gas permeability. J Electrochem Soc 2025;172(6):064507.
[32] Napoli L, Lavorante MJ, Franco J, Sanguinetti A, Fasoli H. Effects on nafion® 117 membrane using different strong acids in various concentrations. J New Mater Electrochem Syst 2013;16(3):151–6.
[33] Schuler Tobias, De Bruycker Ruben, Schmidt Thomas J, Büchi Felix N. Polymer electrolyte water electrolysis: correlating porous transport layer structural properties and performance: Part I. Tomographic analysis of morphology and topology. J Electrochem Soc 2019;166(4):F270–81.
[34] Padgett Elliot, Bender Guido, Alia Shaun M. Membrane pretreatment and cell conditioning for proton exchange membrane water electrolysis. In: ECS meeting abstracts. vol. 2021, 2021, p. 1252.
[35] Alfaki Mustafa M Ali, Masih Shalini Bhawana. Modeling and forecasting by using time series ARIMA models. Int J Eng Res Technol (IJERT) 2015;4(3):914–8.