Energy and AI 21 (2025) 100577

Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.sciencedirect.com/journal/energy-and-ai

Data-driven modeling of polymer electrolyte fuel cells: Towards predictive analytics with explainable artificial intelligence

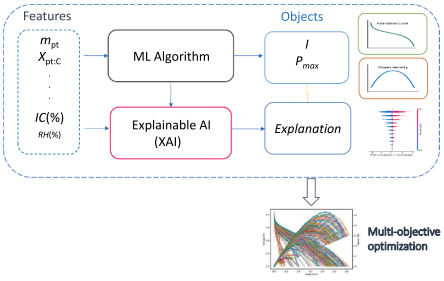
Ali Malek [a], Max Dreger [b], Nima Shaigan [a], Chaojie Song [a], Kourosh Malek [b,*], Jasna Jankovic [c], Michael Eikerling [b,d]

[a] Clean Energy Innovation Research Centre, National Research Council Canada, Vancouver, BC V6T 1W5, Canada
[b] Institute of Energy Technologies (IET): Theory and Computation of Energy Materials (IET-3), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany
[c] Center for Clean Energy Engineering, University of Connecticut, Storrs, CT 06269, United States
[d] Chair of Theory and Computation of Energy Materials, Faculty of Georesources and Materials Engineering, RWTH Aachen University, 52062 Aachen, Germany

HIGHLIGHTS

- Developed a comprehensive data-driven framework using advanced machine learning models to accurately predict polarization behavior of polymer electrolyte fuel cells.
- Integrated explainable AI techniques such as Gini importance or SHAP to reveal how key operational and design parameters especially voltage, relative humidity, platinum loading, and ionomer-to-carbon ratio influence performance.
- Identified voltage is the top predictor, followed by relative humidity, platinum loading, and ionomer-to-carbon ratio.
- Model-based support for optimization of fabrication, real-time control, and durability improvements.
- Providing actionable insights for optimization of the fabrication and operation of polymer electrolyte fuel cells.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Polymer electrolyte fuel cells
Membrane electrode assembly
Machine learning
Explainable artificial intelligence
Performance prediction
Shapley additive explanations

ABSTRACT

Polymer electrolyte fuel cells will be an essential technology of the emerging hydrogen economy. However, optimizing their cost and performance necessitates understanding of how different parameters affect their operation. This optimization problem involves numerous interrelated design and operational parameters. However, developing the required understanding through experimental studies alone would be inefficient. Physical modelling is a much-needed complement to experiment but is constrained by simplifying assumptions that diminish the models' predictive capabilities. As a supplement to experiment and physical modelling, we employ a data-based assessment that leverages machine learning techniques to support and enhance decision-making. We first evaluate the predictive accuracy of various machine learning models, including artificial neural networks, to predict the polarization behavior of polymer electrolyte fuel cells, harnessing an extensive

* Corresponding author.
E-mail address: k.malek@fz-juelich.de (K. Malek).

https://doi.org/10.1016/j.egyai.2025.100577

Available online 29 July 2025

experimental dataset. We then apply explainable artificial intelligence techniques, including Gini feature importance and Shapley additive explanations value analyses, to understand how these models incorporate data into the prediction process. Probabilistic analyses can help identify relationships between predictions and feature values. We demonstrate that insights derived from Shapley additive explanations value analysis are consistent with literature data on the thermodynamics and kinetics of relevant electrochemical reaction and transport processes. Our study highlights the potential of interpretable and explainable tools to offer a holistic analysis of the impacts of various interrelated operational and design parameters on the performance of the fuel cell. In the future, such explainable tools could help identify gaps in experimental data and pinpoint research priorities.

1. Introduction

Climate change has become an omnipresent, undeniable and alarming reality, with environmental and economic implications becoming ever more palpable, although the full extent of long-term consequences is still not fully graspable [1,2]. This situation urges rapid development and adoption of green technologies for power generation and storage. Polymer electrolyte fuel cells (PEFCs) will be an essential part of the emerging global energy economy that will be based on hydrogen as an energy carrier. PEFCs offer high power density, energy conversion efficiency, and durability, rendering them highly suitable for applications in stationary power generation and surface transportation [3].

Despite their long-recognized promise, PEFCs continue to face durability and cost challenges that hinder their large-scale deployment [4]. Current cells struggle to maintain a sufficient level of performance, especially when the amount of precious metal-based electrocatalysts is drastically reduced [5]. Performance and durability of PEFCs are mainly determined by processes in the membrane electrode assembly (MEA), the power-generating unit of a PEFC, which, in turn, depend on various design and operational parameters [4].

The core of the MEA, the so-called catalyst-coated membrane, is a sandwich-like structure consisting of the cathode and anode catalyst layers (CCL and ACL) that are attached to the two faces of a proton-conducting polymer electrolyte membrane (PEM). CL ingredients include nanoparticles of the platinum-based electrocatalyst, a carbon-based catalyst support and finely dispersed ionomer. Mixed together in an ink solution, these ingredients self-assemble into a complex porous composite medium, viz. the catalyst layer, that contains solid regions, polymer regions as well as a porous network comprising a mix of water- and gas-filled pores. Furthermore, the MEA includes so-called gas diffusion layers (GDLs) on both electrode sides that facilitate the supply of reactant gases and as well as the removal of product water.

Many parameters influence the performance and longevity of PEFCs, including the structural properties of components and operating conditions [6,7]. Experimental methods have been used to rationalize relations between structural properties, operating conditions and MEA performance. However, the experimental effort and cost required to comprehend fully the multifarious effects of parameters on MEA behavior would be unaffordable [8,9].

Computational studies and physical modeling have been employed to analyze experimental data, particularly in the modeling of polarization curves, which plot the relationship between cell voltage and current density [10,11]. As a key quantification metric of fuel cell performance, the polarization curve reveals the interplay of electrochemical reactions and transport processes. The analyses of polarization curves is an essential tool to delineate and quantify different contributions to voltage losses due to electrocatalytic reactions, ohmic losses due to limited proton conductivity, and mass transport losses incurred by oxygen transport limitations and the corresponding oxygen depletion.

Different modeling approaches are employed to reproduce, analyse, and, eventually, predict the performance of PEFCs. Modelling efforts encompass empirical models based on mathematical correlations [12–14], semi-empirical models combining empirical data with theoretical insights [15,16], and physics-based analytical models that are based on conservation and transport equations [17–22]. Insights from polarization curve modeling are crucial to identify and understand voltage losses, thus guiding improvements in fuel cell technology.

Variables such as Pt loading, ionomer-to-carbon (I:C) mass ratio, and relative humidity (RH) shape the polarization response of the cell and determine voltage losses incurred by electrochemical reactions, proton transport, reactant gas diffusion, and water transport phenomena. Platinum loading refers to the amount of platinum catalyst in the electrode. A higher Pt loading is beneficial in terms of the electrocatalytic activity, but it also raises the cost of the cell. Muzaffar et al. [20], in their model-based analysis, explored how the drastic reduction of the Pt loading causes unexpectedly large voltage losses. They were able to link this increase to heightened oxygen transport resistances in CCL and particularly the GDL on the cathode side. The I:C ratio, impacts performance and durability via a complex interplay of phenomena. A higher I:C ratio creates beneficial conditions for proton transport and electrochemical reactions at interfaces. However, if the amount of ionomer in the CL is too high, gaseous diffusion of oxygen will be hindered. The RH influences the hydration state (or liquid water saturation) of the PEM as well as the water distribution in porous electrode and diffusion media. Proper hydration ensures high proton conductivity of the PEM. In electrode and diffusion media, insufficient or excessive humidification (i.e., liquid water saturation) affects performance due to dehydration that stifles reactions and proton transport or water flooding that hinders gaseous reactant diffusion.

In addition, operating parameters like temperature and pressure affect reaction kinetics, transport phenomena, and water management, while a large set of intrinsic and effective materials properties (e.g., ion exchange capacity of the ionic polymer, membrane thickness, catalyst layer porosity, proton conductivity, oxygen diffusivity and gas diffusion layer permeability) influence reactant, product, and water transport within the cell [11].

Modeling polarization curves for PEFCs must strike a delicate balance of capturing the most relevant parametric effects and, at the same time, making use of simplifying assumption that keep the mathematical complexity and computational costs involved in solving the model at a reasonable level. Accurately capturing the interplay of multiple parameters is complex, as nonlinear interactions between operating conditions, materials properties, and transport and reaction phenomena render it difficult to isolate individual influences [10]. Furthermore, the computational complexity of numerical simulations of three-dimensional multiphase models, limits their practical application in real-time control and optimization settings for PEFCs [8].

Addressing these challenges, faced by PEFCs, requires advances in materials, enhanced operational control, and design improvements. Artificial intelligence (AI) and machine learning (ML) tools have proven to possess significant potential in addressing some of these limitations in PEFCs [2,23,24]. Their growing role in data analysis, system control, and design optimization is reflected in the increasing number of patents related to these technologies in the energy field [23,24]. They offer new avenues for handling the complexities of PEFC technology, aiding in the development of fundamental knowledge, material selection, fuel cell design, and optimization. AI and ML are effective in managing system control, power management, and operational health monitoring [24–26]. Among ML models, supervised learning and deep learning (DL)

techniques, have shown to be particularly valuable for analyzing PEFC polarization behavior [27–29]. Trained on sufficiently large experimental datasets, ML can predict performance outcomes under various conditions with high accuracy, surpassing traditional modeling techniques that often require impractically large computational resources. Moreover, ML can facilitate the optimization of MEA properties by predicting how changes in material composition or operating conditions affect the overall efficiency and longevity of the fuel cells.

Recent advances in AI and ML have the makings of revolutionizing PEFC and electrolyzer research, as they could enable more accurate prognostics and remaining useful life (RUL) predictions through hybrid physics-based and data-driven models that even account for voltage recovery [30]. These approaches, often enhanced with explainable AI (XAI) techniques like Shapley additive explanations (SHAP), provide essential insights into impacts of design and operational parameters and pathways to the optimization of performance and lifetime [29,31]. The emerging data-driven paradigm will transform the manufacturing processes and advance clean energy technologies by moving beyond "black box" predictions to offer mechanistic understanding [32].

For example, Wang et al. [33] introduced an AI-based framework that uses surrogate modelling and optimization to improve the maximum power density of PEFCs by optimizing the catalyst layer (CL) composition. The framework employs a data-driven surrogate model trained with a database generated from a computational fluid dynamics (CFD) model, and it demonstrates high accuracy and computational efficiency. Rui et al. [34] proposed an ML methodology called SPARK (Smart Prediction of Advanced Research on PEMs using Knowledge-based machine learning), which predicts the performance and durability of PEMs in fuel cell applications. By utilizing the SPARK methodology, the design process for ceria-containing PEMs was streamlined, reducing experimental costs and development time while enabling efficient design with specific characteristics. Recently Ding et al. [35] developed visualizable algorithms to aid researchers in decision-making to uncover the effect of important MEA parameters. They developed a comprehensive set of ML-assisted processes, consisting of four modules (feature selection, decision modelling, regression modelling and extremum optimization), to address limitations, such as inefficient data utilization, lack of comprehensive approaches, and lower prediction accuracy, and to achieve high accuracy in predicting key performance parameters. Current data-driven approaches have been primarily used for making accurate predictions of performance and durability , and the operando monitoring of fuel cells. ML techniques are utilized in material selection, chemical reaction modeling, and polarization curve analyses. For durability, ML aids in diagnosing faults, estimating the state of health, and predicting the remaining useful life [36]. However, ML models often lack transparency in their decision-making processes. These approaches are not capable of explaining how different parameters affect the model's prediction. Global and local interpretability and explainability tools may be applied to ML models for interpretation [37]. In general, explainable ML tools can be divided into global and local explainable methods. Global importance methods assess the overall significance of features across the entire model, providing a holistic view of feature contributions [31,38,39]. Conversely, local importance methods, such as SHAP values [40], allow evaluating the impact of individual features on the prediction of a given target.

To address the limitations of current data-driven approaches, we propose a data-driven framework that harnesses ML methods to accurately predict the polarization response of PEFCs as function of RH, platinum loading, platinum-to-carbon ratio (Pt:C) and ECSA. Unlike conventional approaches that treat ML models as "black boxes", our framework integrates a novel combination of both global (e.g., Gini feature importance) [38,39] and local (e.g., SHAP values) [40] interpretability techniques to provide fundamental understanding on the effect of parameters on the polarization behavior. This integrated methodology enables mechanistic insight into how specific individual parameters influence PEFC performance, thereby bridging the gap between predictive power and scientific fundamental understanding.

Crafting this framework to a point that it able to deliver accurate and reliable predictions demanded an extensive effort, including the curation of a vast and diverse dataset, rigorous model training, and validation across diverse array of operating conditions. This is essential for an efficient and fast optimization process as is needed to facilitate the industrialization of PEFC technology. In addition, the framework developed can be readily adapted and applied to other electrochemical technologies, including optimization of the emerging PEM water electrolysers.

2. Methodology and model implementation

2.1. DATA sources and considered features

In this study, datasets of MEA polarization curves were collected at the Fraunhofer ISE in Freiburg, Germany, through standard experimental methods, using a conventional single-cell configuration [41]. They fabricated MEAs with various Pt loadings and I:C ratios using a screen-printing process. A Baltic FuelCells quickConnect® "high amp" test cell with straight flow channels and an active area of 3×4 cm^2 was employed. MEAs were fabricated using Gore membrane M735.18, and cathode catalysts of 40 wt. % and 60 wt. % Pt:C. Pt loading was varied from 0.05 to 0.8 mg/cm^2, and I:C ratios ranged between 0.29 and 1.67. The cell was operated at 80 °C with fully humidified reactant gases and back pressures up to 2.0 bar.

MEAs underwent extensive in situ analyses within a test cell, which included polarization curve measurements to evaluate voltage performance under different operational conditions, cyclic voltammetry to determine the electrochemical surface area (ECSA), and humidity sweeps to assess the effects of humidity on cell performance. These methods provide comprehensive insights into the impact of alterations in catalyst layer composition on overall functionality and performance.

Fig. 1(a) and (b) depicts a subset of PEFC polarization curves (see Figure S.1 in supporting information for the full set), displaying output voltage and power density at RH of 40 % and 100 %. Table 1 lists the ranges of parameters that are used to train the ML models for the prediction of the MEA performance. In the dataset, RH was the only operating parameter that was adjustable to enhance performance during the PEFC operation. Design parameters extracted from the database include Pt loading (mg/cm^2), Pt:C ratio and I:C ratio.

To ensure the accuracy and reliability of the data-based modelling, we took precautions to actively prevent data leakage. Data leakage is a common issue that can lead to overly optimistic performance estimates and degrade model reliability on new or previously unseen data. It occurs when the training process uses features that are not available at prediction time. In our study, properties such as maximum power density, which is derived from predicted polarization curves, could cause such leakage. To mitigate this risk, we excluded maximum power density and similar derived variables from the input feature set. Instead, the parameters used as input features have been restricted to fabrication and compositional parameters along with operational parameters which are the primary parameters determining the shapes of polarization curves. By excluding dependent parameters such as maximum power density, we were able to enhance the robustness and relevance of our models, ensuring their applicability to real-world PEFC scenarios.

It is worth emphasizing that we have used voltage as an input feature to predict the current density in PEFC polarization curves. Using voltage as input feature in data-based modeling of PEFCs offers several advantages. It enables inverse modeling, which allows for versatile performance analyses and predictions of current density at various voltage points without the need for data model retraining. This enhances the model's versatility, potentially supporting both forward (voltage prediction) and inverse (current density prediction) applications, which is valuable in different analyses. Furthermore, considering voltage as an

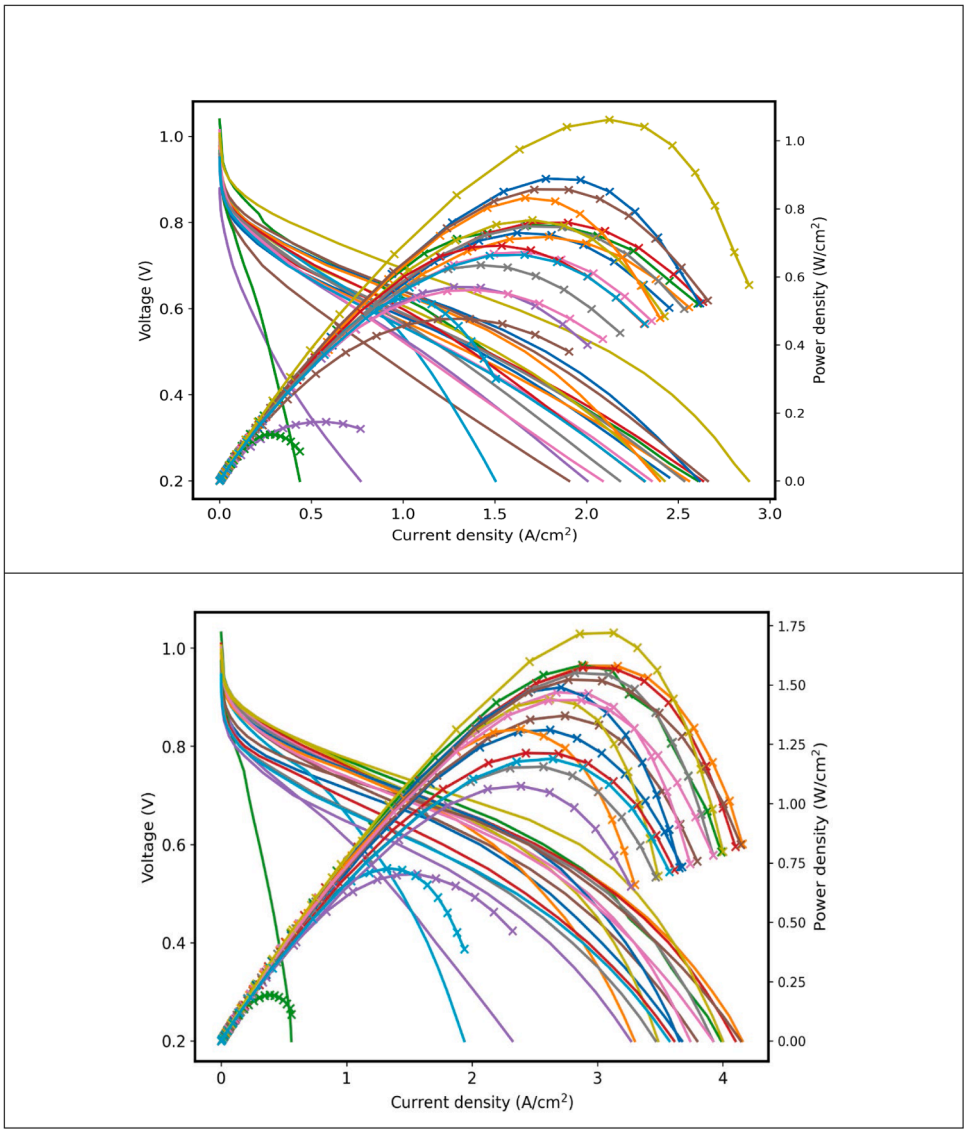


Fig. 1. Polarization and power density of PEFCs under different relative humidity conditions.
(a) Polarization curves measured at 40 % for a selection of 20 representative cells.
(b) Corresponding power density curves for the same cells tested at 100 % relative humidity, illustrating the effect of relative humidity on performance.

Table 1
PEFC performance variables used in machine learning models.

Feature	Lower bound	Upper bound
Measurement ID	1	145
Pt Loading [mg/cm^2]	0.055	0.78
Ionomer [wt %]	15	50
Pt:C Ratio [%]	40	60
RH [%]	40	100
ECSA [cm^2/cm^2]	2.0	298
Voltage [V]	0.19	1.4
Current density [A/cm^2]	0	4.55

input feature allows ML models to more effectively capture the nonlinear relationship between voltage and current density.

The preprocessing of the MEA performance dataset demanded several steps to ensure its suitability for ML and DL analyses. The dataset underwent a thorough check for missing values and inconsistencies, with anomalies addressed to ensure data integrity. Duplicated data entries were removed by identifying identical feature values. Outliers were detected using the interquartile range (IQR), which defines outliers as data points lying below Q1 − 1.5 × IQR or above Q3 + 1.5 × IQR, where Q1 and Q3 are the first and third quartiles, respectively. These outliers were handled appropriately to prevent skewness in the model training process. Furthermore, each feature was standardized to achieve zero mean and unit variance, crucial for minimizing biases in ML models, particularly those sensitive to input data scales like support vector machines (SVM) [42] and neural networks (NN).

2.2. *Machine learning models*

A diverse array of ML models was employed to ensure robustness, reduce bias, enhance predictive accuracy and to benchmark and compare their performance. The models included tree-based algorithms such as XGBoost [43] and Random Forest (RF) [44], alongside linear models like standard linear regression (LR), Ridge [45] and Lasso Regression [46] and advanced boosting algorithms like LightGBM [47], and CatBoost [48]. This diverse model selection ensures that both linear and non-linear relationships are effectively captured. Ensemble methods help reduce overfitting, while regularization techniques in linear models prevent the escalation of model complexity. Diversity of the models

allows for efficient management of different data characteristics, e.g., multicollinearity and complex interaction effects among variables. The dataset was divided into training and testing sets (80 % training and 20 % testing) and controlled by a fixed random seed to ensure reproducibility. Hyperparameter tuning was conducted through 5-fold GridSearchCV during the training phase. This enables a systematic exploration of parameter combinations based on cross-validation performance [54].

ANN and Deep Belief Networks (DBN) [49] (see supporting information for more details) with TensorFlow were utilized to further enhance the prediction accuracy and account for specific data characteristics. The training dataset was augmented to reinforce the models' accuracy and robustness by fitting and interpolating and generating additional data points. Data imputation improves the training quality of DL models by incorporating a broader range of inputs. This improves the reliability of data-based analytics and reduces computational efforts.

Performance metrics were extracted for both training and test datasets to evaluate the predictive accuracy for ML and DL models. These metrics included root mean square error (RMSE), normalized root mean square error (nRMSE), mean absolute error (MAE), coefficient of determination (R^2), and Pearson correlation coefficient [50]. RMSE and MAE determine the magnitude of prediction errors. Values approaching 0 indicate the greatest accuracy. nRMSE provides an error measurement relative to the mean actual value and enhances the interpretability across different datasets. R^2 values are a measure of the model's consistency. R^2 values approaching 1 indicate the greatest consistency. Pearson correlation assesses the linear relationship between observed and predicted values.

2.3. EXPLAINABLE machine learning

While ML models are powerful, they often lack interpretability and explainability [51]. Gini feature importance [52] and SHAP techniques [40] are commonly used to explain ML models. Gini feature importance, derived from the Gini impurity, quantifies a feature's importance by measuring the total decrease in node impurity, averaged over all trees in the forest. Features that result in larger decreases in Gini impurity are considered more important, as they contribute more to the model's prediction. Despite its computational efficiency, Gini importance can be biased toward high-cardinality features and may not capture complex feature interactions. On the other hand, SHAP value analysis, based on Shapley values from game theory, is used to explain the ML models both locally and globally. SHAP value analysis determines the contribution of each feature to the model's prediction by treating feature values as players in a coalition game [53].

At the local level, features with positive SHAP values have a positive impact on predictions, while those with negative values have a negative impact. Distribution profiles show how a certain feature influences the model's prediction. The magnitude of SHAP values quantifies the maximum and minimum extents to which a feature can alter the predicted values. Overall, SHAP values provide a robust framework for interpreting the model's behavior, identifying influential features, and enhancing the transparency of ML predictions.

This model-agnostic approach can be applied to various models, including linear regression, decision trees, RFs, gradient boosting models, and NNs. Local SHAP values explain a specific prediction by attributing contributions to individual features. It can be represented as:

$$g(\mathbf{z}') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j \tag{1}$$

where $g(\mathbf{z}')$ represents the validated ML model output for a simplified input $\mathbf{z}'$, ϕ_0 is the base value (i.e., the expected model output), $z'_j \in \{0, 1\}$ indicates whether feature j is present or absent in the coalition vector z', ϕ_j is the SHAP value for feature j, and M denotes the number of features. Therefore, a higher magnitude of ϕ_j shows a greater contribution of feature j to the prediction compared to others.

While individual SHAP values provide local explanations for specific predictions, the global SHAP feature importance analysis aggregates these local values across multiple instances accordingly. The aggregation is defined as:

$$I_j = \frac{1}{N} \sum_{i=1}^{N} \left| \phi_j^{(i)} \right| \tag{2}$$

where I_j denotes the global importance of feature j, calculated as the average absolute SHAP values across all instances, $\phi_j^{(i)}$ is the SHAP value for feature j in instance i, and N is the total number of samples in the dataset. Features with higher I_j values are ranked accordingly.

SHAP values possess key properties including additivity, local accuracy, missingness and consistency. Additivity allows for the independent computation and summation of feature contributions. Local accuracy ensures that SHAP values match the difference between expected and actual model outputs for given inputs. The missingness property assigns a zero value to features that are irrelevant or missing, thereby preventing distortion. Lastly, consistency ensures SHAP values remain unchanged unless there is a change in contributions of different features.

Beeswarm SHAP plots are a powerful and informative visualization tool for SHAP value analysis. They offer a comprehensive view of how various features influence the model's decision-making process. A Beeswarm plot ranks features vertically according to their importance for the model's predictions based on their mean absolute SHAP values across the entire dataset. The most and least important features appear at the top and bottom of the plot, respectively. For each feature, the data points (i.e., dots) represent individual instances in the dataset (i.e., a row in the dataset).

These data points are distributed horizontally along the SHAP value axis, reflecting a certain range of variation in their contributions to the predicted outcome. Feature importance is ranked based on the extent of horizontal distribution. Distribution profiles may appear as a narrow band (e.g., 1–2 dots thick) or as clusters of vertically stacked dots, representing datapoints. A narrow distribution profile indicates that the feature has a strong and consistent influence on the model's predictions. This suggests the model sees the data relevant in this narrow range and uses them for prediction. If the predicted values are a function of a specific feature, the distribution profile for that feature will appear as a narrow line. Clustering and vertical stacking of datapoints with the same SHAP value, on the other hand, illustrate an insignificant effect on the predictive capability of the model. This may be due to a high degree of scatter in the raw data (i.e., high standard deviation of values) or the lack of a clear relationship between the feature and predicted values. In such cases, the model does not tend to use features' values, especially those that show clustering or a vertically stacked distribution profile within a specific range when making prediction.

Feature values for each instance (distinct from their SHAP values) are represented through a color scale. The highest and lowest values for each feature appear as bright red and blue dots, respectively, on the color bar of SHAP value plots. Intermediate values appear as different shades of violet. This color mapping helps distinguish the effect of low versus high feature values on the model's prediction. For example, if the narrow distribution profile is composed mainly of blue dots within a certain SHAP value range, it implies that the model uses the lower values of that feature to predict the outcome. The SHAP value axis is divided into negative and positive regions. If a feature has a positive impact on the outcome, the datapoints appear on the positive side. Conversely, the negative side of the axis contains the datapoints for features that negatively impact the outcome. The extent of a feature's impact, positive or negative, is indicated by the magnitude of its horizontal spread along the SHAP axis.

3. Results and discussions

3.1. Evaluation of models' prediction accuracy

Fig. 2 (a-h) illustrates the comparison between the measured current density values and values predicted by different ML algorithms and ANN and DBN methods. The linear behavior of these plots indicates a high correlation between predicted and experimental current density values.

Support Vector Regression (SVR), RF, Bagging, CatBoost, XGBoost, and Histogram-based Gradient Boosting (Hist Gradient Boosting) [55] demonstrated higher predictive accuracy than other models studied. The R^2 values are listed in Table 2. ANN and DBN models effectively capture the non-linear relationships and complex interactions among input features. This is important given the dynamic nature of PEFC operation.

Figs. 3(a) and 3(b) show the predictive performance of the models based on RMSE and nRMSE values, respectively. The mean RMSE across all models was 0.23 ± 0.09. The mean RMSE represents the average deviation of the predicted and actual current density values as a direct measure of prediction error. To enable a meaningful comparison among the models, we normalized the RMSE (resulting in nRMSE) by the interquartile range (IQR) of the actual current density values from the experimental dataset. This normalization reduces the impact of outliers. The mean nRMSE value is 0.097 ± 0.049.

Table 2 compares the performance metrics of all ML models (see Table S.1 for training data set and optimal model parameters in supporting information). Among the models studied, SVR exhibits the lowest RMSE (0.12), nRMSE (0.06) and the highest R^2 value (0.99). These values indicate better predictive accuracy and minimum error for SVR. SVR also ranks high in terms of correlation, indicating that the model's ability in effectively capturing the underlying relationships between the features and outputs.

The RMSE and nRMSE values for ensemble methods, such as RF and Bagging were 0.12 and 0.06, and 0.13 and 0.06, respectively. The predictive accuracy of RF and Bagging closely matches or exceeds that of SVR based methods, $R^2 = 0.99$. The high prediction accuracy of the ensemble techniques can be attributed to their efficacy for modeling complex data structures.

The RMSE, nRMSE and R^2 values for the ANN model were 0.13, 0.061 and 0.99, respectively (see Table S.2). The prediction accuracy for the ANN model, as estimated by the individual performance metrics, does not seem to be great. However, it ranks high when all performance metrics are taken into account. This may be attributed to ANN's balanced performance for modeling non-linear relationships.

The RMSE, nRMSE and R^2 values for the DBN model were 0.16, 0.078 and 0.98, respectively. Both, ANN and DBN models, have correlation scores of 0.98. Therefore, ANN and DBN are efficient models in terms of data fitting and uncovering the relationships between features and predicted values. Effective handling of feature interactions, robustness against overfitting, and efficient learning of complex patterns are the key characteristics of ANN and DBN models that make them suitable for modeling the fuel cell polarization behavior. Specifically, ANN and DBN benefit from their DL capabilities, which is necessary for understanding subtle, complex and multilayered interactions among input features.

Both CatBoost and XGBoost, as gradient boosting models, achieved an RMSE of 0.15 and an R^2 of 0.985, indicating comparable and high predictive accuracy. These models are, therefore, suitable for modeling complex predictive scenarios. The RMSE of 0.17 and an R^2 of 0.98 for the histogram-based Gradient Boosting model indicate the model's capacity for addressing complex dataset features.

Linear models including Linear Regression, Ridge, Lasso, and Lars, showed similarly high RMSE and nRMSE values of approximately 0.40 and 0.19, respectively. The R^2 value for the linear models was approximately 0.9, smaller than that for the other models. Since the polarization curves are only linear in the intermediate regime, linear models fail to predict the entire range of polarization data. Among the linear models, Lasso showed the greatest RMSE value of 0.406, implying the method cannot effectively select features in the presence of multiple correlated parameters, leading to underfitting. The other models studied, on the other hand, showed unacceptable prediction accuracy. AdaBoost, despite its high correlation score, showed the highest RMSE and nRMSE values, 0.31 and 0.15, respectively. These values indicate substantial average prediction errors. The model's large MAE value of 0.26 suggests that AdaBoost is not suitable for outlier predictions and modeling complex, non-linear data complexities.

Due to their superior predictive accuracy and reliability, SVR, RF, Bagging, XGBoost, CatBoost, and Hist Gradient Boosting were selected for feature importance analysis. These models employ ensemble and boosting techniques to effectively capture non-linear relationships and incorporate regularization to avoid overfitting.

Linear models, such as Linear Regression, Ridge, Lasso, and Lars, showed similarly high RMSE and nRMSE values of approximately 0.40 and 0.19, respectively. The R^2 value for the linear models was approximately 0.90, lower than the other models. These models assume linear relationships between variables, limiting their ability to handle the non-linear complexities typically present in PEFC data. Among the linear models, Lasso showed the highest RMSE value (0.41), indicating its disadvantage in feature selection in the presence of multiple correlated variables leading to underfitting.

The top performing ML models, SVR, RF, Bagging, XGBoost, CatBoost, and Hist Gradient Boosting, along with ANN and DBN, can greatly enhance the predictive accuracy and reliability in PEFC systems. These models can capture non-linear relationships, employ ensemble and boosting techniques, and incorporate regularization to avoid overfitting.

Specifically, ANN and DBN are notable for their DL capabilities, which allow for a nuanced understanding of complex, multilayered interactions among input features. Their high R^2 (ANN: 0.99, DBN: 0.98) and correlation scores (ANN: 0.98, DBN: 0.98) demonstrate their good performance in accurately fitting data and revealing underlying relationships between input features and output current density.

Consequently, their sophisticated handling of feature interactions, robustness against overfitting, and efficient learning of complex patterns make them invaluable for advancing PEFC applications in the energy sector.

3.2. Explainable machine learning: feature importance

Fig. 4 (a-d) shows the Gini feature importance ratings, extracted for the top-performing models including RF, CatBoost, XGBoost, and LightGBM. The Gini importance, a measure of each feature's contribution to the model's predictive power, consistently identifies voltage as the most significant feature across all models except LightGBM. Pt loading and RH show varying levels of importance across different models.

RF shows a balanced consideration of all features, except ECSA, indicating its efficacy for leveraging a broad spectrum of interdependent inputs. For LightGBM, platinum loading was the most significant feature. This model uses a gradient-based learning technique, focusing on feature interactions. For CatBoost and XG Boost models, voltage and RH are the most important features. The algorithms used for these models are effective for processing categorical and continuous interactions among features.

RH showed a greater importance for XGBoost and LightGBM. For these models, I:C and Pt:C ratios were less significant than voltage and platinum loading. However, I:C ratio and Pt:C ratio have a great impact for RF and CatBoost. ECSA appears to be less important for most models, except LightGBM. The variation in feature importance may suggest that an ensemble approach, e.g., integrating insights from multiple models, may be effective for capturing a broader spectrum of features.

3.3. Explainable machine learning: SHAP values

Fig. 5 (a-d) presents the Beeswarm SHAP value summary plots for the

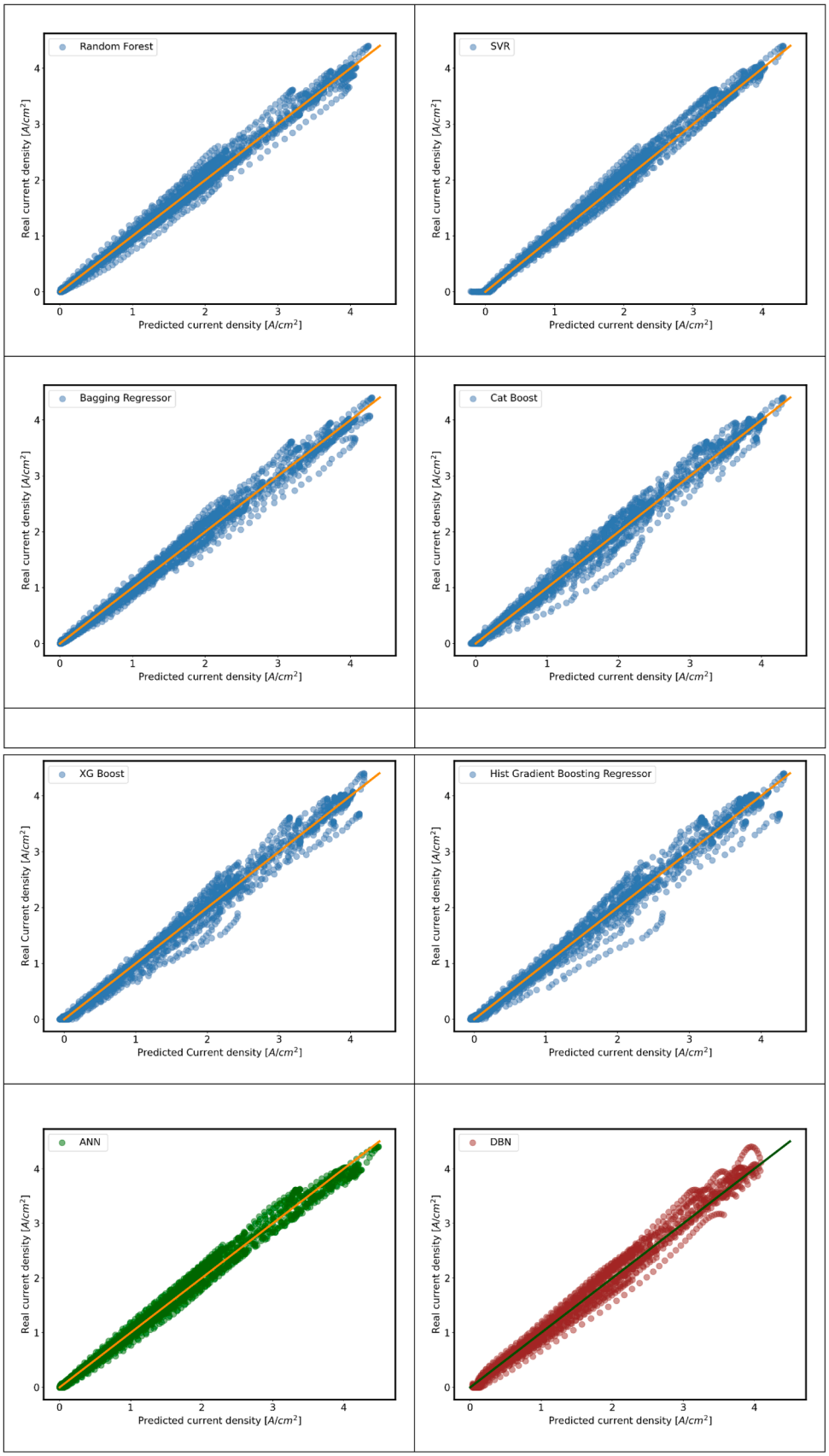


Fig. 2. Correlation between real and predicted current density (A/cm^2) values PEFCs using various machine learning models: a) Random Forest (RF), b) Support Vector Regression (SVR), c) CatBoost, d) Bagging, e) Histogram-based Gradient Boosting (Hist), f) Extreme Gradient Boosting (XG Boost), g) Artificial Neural Network (ANN) and h) Deep Belief Network (DBN).

Table 2
Comparative performance metrics of various machine learning models for predicting current density in PEFCs.

Model	Test RMSE	Test nRMSE	Test MAE	Test R^2	Test Correlation
XGBoost	0.154	0.073	0.105	0.985	0.978
LightGBM	0.195	0.092	0.127	0.977	0.953
Gradient Boost	0.203	0.096	0.145	0.975	0.969
Random Forest	0.125	0.059	0.086	0.990	0.978
Extra Tree	0.170	0.080	0.109	0.982	0.972
Decision Tree	0.170	0.080	0.109	0.982	0.972
AdaBoost	0.315	0.148	0.264	0.939	0.984
CatBoost	0.154	0.073	0.106	0.985	0.974
Linear Regression	0.401	0.189	0.320	0.901	0.957
Ridge	0.401	0.189	0.320	0.901	0.957
Lasso	0.406	0.191	0.327	0.899	0.957
Lars	0.401	0.188	0.323	0.901	0.954
Bagging	0.130	0.061	0.089	0.990	0.978
SVR	0.118	0.056	0.087	0.991	0.979
Hist Gradient Boosting	0.168	0.079	0.107	0.983	0.971
Stacking	0.174	0.082	0.114	0.981	0.976
ANN	0.129	0.061	0.093	0.989	0.979
DBN	0.165	0.077	0.130	0.981	0.982

top performing models, including SVR, RF, Bagging Regressor, and CatBoost. There is a strong correlation observed between the SHAP value and Gini feature importance analysis. Both methods indicate that voltage has the greatest influence on all models' predictions. RH ranks as the second most important feature. While I:C and Pt:C ratios are important across all models, their rankings differ slightly between the SVR and other models. Pt loading is ranked higher in importance for the RF model compared to the SVR model. ECSA is identified as the least important feature across all models except SVR, where ECSA's contribution to the overall prediction of current density is negligible. The consistency observed across different explanatory methods underscores the reliability of interpretations of feature importance.

Voltage: Voltage has a similar distribution in the SHAP plots for all models. Generally, the voltage of an electrochemical cell, V_{cell}, is the equilibrium cell potential E^0_{cell} minus the sum of all overpotentials:

$$V_{cell} = E^0_{cell} - \eta_{act} - \eta_{ohmic} - \eta_{mass} \quad (3)$$

where η_{act}, η_{ohmic} and η_{mass} represent the activation, ohmic and mass transport overpotentials, respectively. Generally, polarization curves for electrochemical cells show three different regimes including activation regime, intermediate (or ohmic) regime and oxygen depletion regime, as rationalized and explained in Refs. [11,22].

As the plots in Fig. 5 (a-h) show, close to equilibrium or open circuit voltage the current density prediction from the voltage shows the highest level of uncertainty, caused by inaccuracies in voltage measurements and the high sensitivity of the current to the voltage in this region [56,57]. By moving farther from left to right on the horizontal axis, the dots representing voltage converge.

In the kinetic regime, the predicted current density is exclusively a function of voltage, independent of other features. The activation overpotential, relevant in this regime, is a function of operating temperature, ECSA, exchange current density, j^0, and charge transfer coefficient, α. Resistance contributions from ionic and electronic conduction minimally affect the voltage, and can be ignored in the kinetic regime. However, for this assumption to be true, values of I:C and Pt:C ratios must exceed a critical threshold. Specifically, the I:C ratio should be greater than 20 %, and the Pt loading should be >0.2 mg/cm^2.

Moving farther to the right on the horizontal axis, the degree of vertical stacking increases again. This region coincides with the onset of the intermediate regime in polarization curves. The vertical scattering of dots indicates that the model uses features other than voltage to predict current density. Voltage in the intermediate region is a function of ohmic resistances, primarily incurred by the proton transport. In the intermediate regime, voltage is strongly dependent on the protonic conductivity and the thickness of the CCL. The proton conductivity, in turn, is a function of the liquid saturation, which is influenced by local RH, pore size distribution, wettability, as well as the dispersion and microstructure of the ionomer [58]. The electronic conductivity is a function of the carbon loading (and the type of carbon), Pt:C ratio, I:C ratio and CL thickness. As the current increases, datapoints converge in the middle of the intermediate region and diverge again into the oxygen depletion region.

A cluster of blue dots appears at the end of the positive SHAP value axis between 1.2 A/cm^2 and 1.8 A/cm^2 (SHAP values for SVR in Fig. 5 (b)). This cluster represents the oxygen-depletion regime of polarization curves. The high fluxes of oxygen and water required in this regime may cause, in addition to significant voltage losses, reversible or irreversible structural changes in the transport paths, which depend on the prevailing local conditions and the transient protocol of operating states the sample has been subdued to. In other words, the repeatability of experiments in the oxygen depletion regime is lower compared to that in the intermediate region.

Relative humidity: The SHAP value distribution profile for the RH shows two distinct unicolor regions, either blue or red corresponding to 40 % and 100 % RH, respectively. Switching the RH from 100 % to 40 % has a negative effect on the predicted current density. The minimum and maximum of the predicted current density for the SVR model shift to more negative values by approximately 0.15 A and 0.66 A, respectively. Positive and negative effects are virtually symmetrical. Other models show similar SHAP value profiles for RH.

The proton transport resistance in the MEA decreases as the RH increases. Therefore, the effect of RH is expected to be the most pronounced in the intermediate regime of the polarization curve. Vertical stacking and clustering of dots, observed at both ends of RH bands, correspond to the kinetic and oxygen depletion regimes. Vertical stacking is more pronounced for the oxygen depletion regime than that for the kinetic regime. In the kinetic regime, the current density is small and the proton transport resistance is not a controlling factor. Therefore, upon increasing the RH the current density remains stable in the kinetic regime, as long as the liquid saturation and the ECSA remain relatively unaffected by the RH. The vertical stacking of dots in the oxygen depletion regime indicates that the predicted current density is not significantly affected by the RH, as increased RH (and thus an increased liquid saturation) will have negligible or negative effect on mass transport. At RH of 100 %, the probability of saturation and flooding increases, exacerbating mass transport limitations.

Ionomer to carbon mass ratio (I:C ratio): In the positive region of SHAP values, in Fig. 5 (a-d), vertical stacking of blue and red dots indicates that ionomer content is not considered as an important feature taken into account by the model. In the negative region of SHAP values, for I:C ratio in the range of 15 and 20 % (indicated by bright blue dots), a negative effect on the predicted current density is observed. The color shifting from lighter to darker blue (moving in the positive direction on SHAP value axis) implies that when I:C ratio < 25 %, the predicted current density decreases by 0.3 A/cm^2.

Vertical stacking at larger I:C ratio (dark blue dots) reveals that the current density is independent of I:C ratio, when this ratio exceeds a certain threshold. This suggests that future research should focus on determining the I:C ratio threshold value. Levels of the I:C ratio that exceed an optimal value (here the optimal value is expected to lie in the range around 30 %) may have an adverse effect on cell performance. An excessive amount of ionomer can increase the CL thickness, reduce the porosity of the CL and increase the wettability (and, thus, propensity for flooding) of the pore space that is needed for gaseous oxygen supply [58], pushing the polarization curve into the oxygen depletion region at smaller current density.

Pt:C (mass) ratio: The dataset used in this work contained two, high and low values for Pt:C ratio, 40 % and 60 %. In Fig. 5 (a-d), vertical

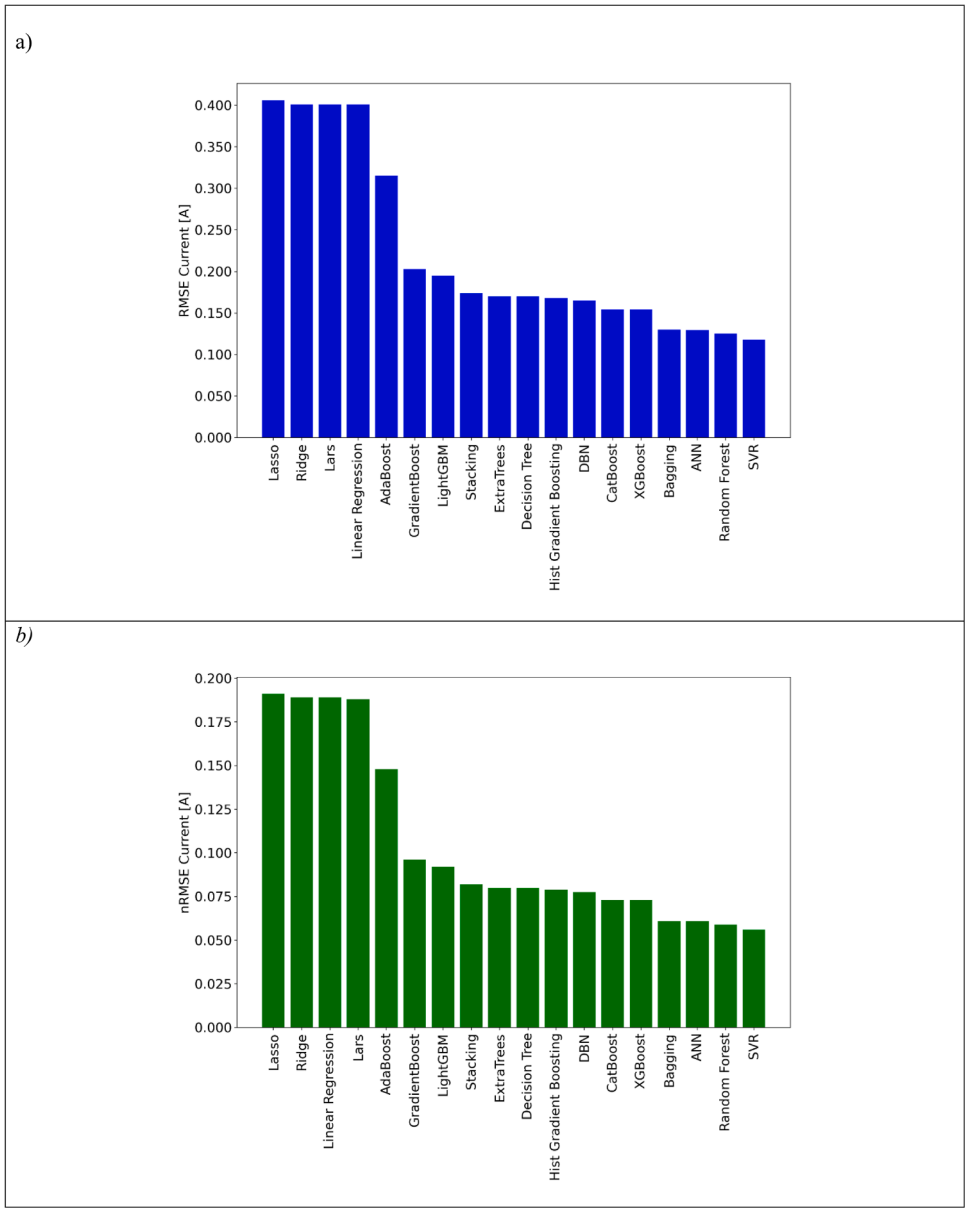


Fig. 3. a) Root Mean Squared Error (RMSE), and b) Normalized Root Mean Squared Error (nRMSE) for predicted current density values using various machine learning models.

stacking of blue points in the summary plot for Pt:C ratio of 40 %, suggests that the data associated with the lower value for this feature were not considered for all model calculations. Instead, the model focuses on data with Pt:C ratio of 60 % for prediction. The standard deviation of data corresponding to Pt:C ratio of 60 % is smaller than that for Pt:C ratio of 40 %.

The Pt:C ratio has a great influence on CL thickness for a given value of platinum loading, due to a large difference between platinum and carbon densities [41] . Increasing the Pt:C ratio from 40 % to 60 % decreases the CL thickness by 33 %. Reducing the CL thickness leads to an increase in proton conductance (and electron conductance). Likewise, the pathways for oxygen diffusion become shorter with decreasing CL thickness, which is beneficial. At Pt:C ratios of 60 % and 40 %, depending on other features, predicted current density values may increase by 0.3 A/cm^2 in the voltage range between 0.6 and 0.75 V

Platinum loading: The SHAP summary plots for different models in Fig. 5 (a-d) show a similar effect for platinum loading. The negative regions of SHAP value plots for Bagging Regressor and CatBoost reveal a greater decrease in predicted current density for lower values of platinum loading in comparison with those for SVR and RFs. The maximum decrease in predicted current density values for minimum platinum loading was approximately 0.55 A/cm^2 for Bagging Regressor and CatBoost models.

Distribution of SHAP values suggest that the lower values of Pt loading (blue dots) negatively affect the predicted current density in kinetic and intermediate regimes. Higher values (red dots) of platinum loading show clustering, indicating that the model does not rely on the higher values of platinum loading. A minimum threshold value for Pt loading is required to achieve an acceptable performance. Greater values will make no difference for the current density prediction. This is consistent with the threshold value reported in the work by Muzaffar *et al.*, which was found in the range of 0.05–0.1 mg cm^{-2} [20]. In the oxygen depletion regime, the current density is independent of Pt loading as indicated by clustering of blue and red dots. Increasing Pt loading causes an increase in CL thickness, which let oxygen transport losses grow [20,41]. The reduction in kinetic losses resulting from increased Pt

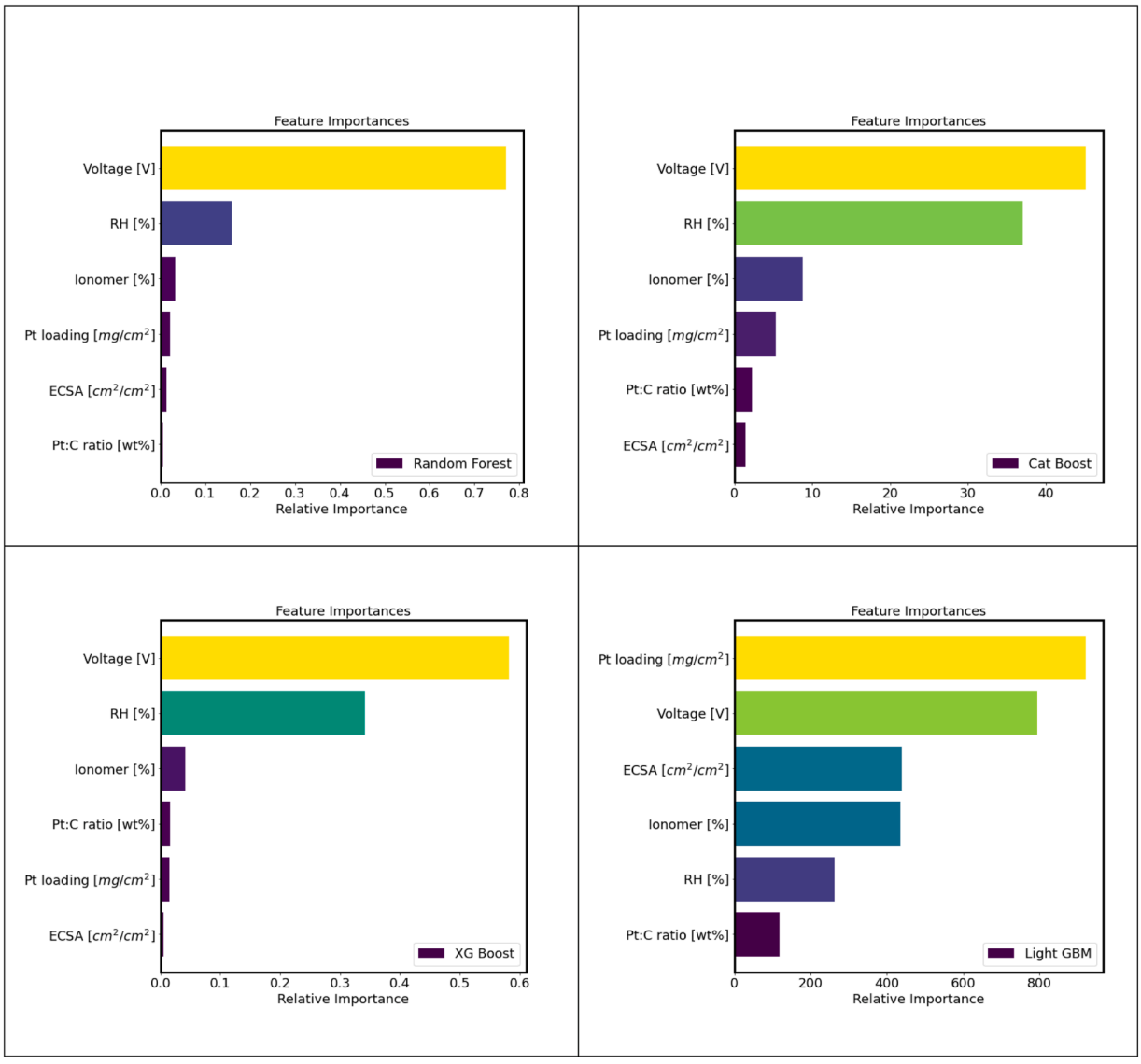


Fig. 4. Comparative Gini feature importance values from various machine learning models used for predicting current density in PEFCs. (a) Random Forest, (b) CatBoost, (c) XGBoost, and (d) LightGBM.

loading is compensated by the increase of ohmic transport losses in the intermediate regime and overcompensated by the transport losses (due to oxygen and protons combined) in the oxygen depletion regime.

ECSA: For ECSA, the SHAP summary plots in Fig. 5 (a-d) show points densely clustered around zero SHAP values. Irrespective of variations in ECSA values, the impact on the model's prediction of the current density is negligible. ECSA is a complex function of volume fractions of Pt, Pt:C ratio, I:C ratio, and porosity in the CL.

Overall, as demonstrated with this analysis and discussion, explainable AI methods can provide valuable information on how different features affect predicted current density values in different regimes of PEFC polarization curves. The foregoing discussion exemplifies, how this analysis could single out certain features as important factors in the prediction (and thus design) process; however, this discussion is by no means meant to be comprehensive in this regard. Our discussion reveals that the feature sensitivity varies significantly, depending on the regime of a polarization curve that is considered.

If the interpretations reached on the basis of explainable AI methods are consistent with independent experimental findings, valuable conclusions about mechanistic or parametric effects may be drawn from them. Therefore, the application of such methods will be able to yield recommendations for a more focused and more efficient research process in designing and optimizing MEAs for PEFCs.

4. Conclusion

This study has demonstrated a robust data-driven framework for modeling of PEFC polarization behaviour, using various advanced ML techniques including ensemble methods such as RF, CatBoost and ANN and DBN models. These models achieved exceptional predictive accuracy, with R^2 values nearing 0.99 and a low RMSE, effectively capturing the non-linear dynamics of PEFC operation and surpassing the limitations of traditional linear and physics-based models.

By integrating explainable AI tools, such as Gini feature importance and SHAP value analysis, the study provides deep mechanistic insights into PEFC performance. SHAP value analysis statistically qualifies and quantifies the effect of different features and their value ranges on the model's predicted current density values.

The results of the detailed feature analysis were interpreted in terms of the physical phenomena occurring in an MEA in different regimes of polarization curves. Based on SHAP analysis and Gini feature importance results, voltage was identified as the most influential feature in

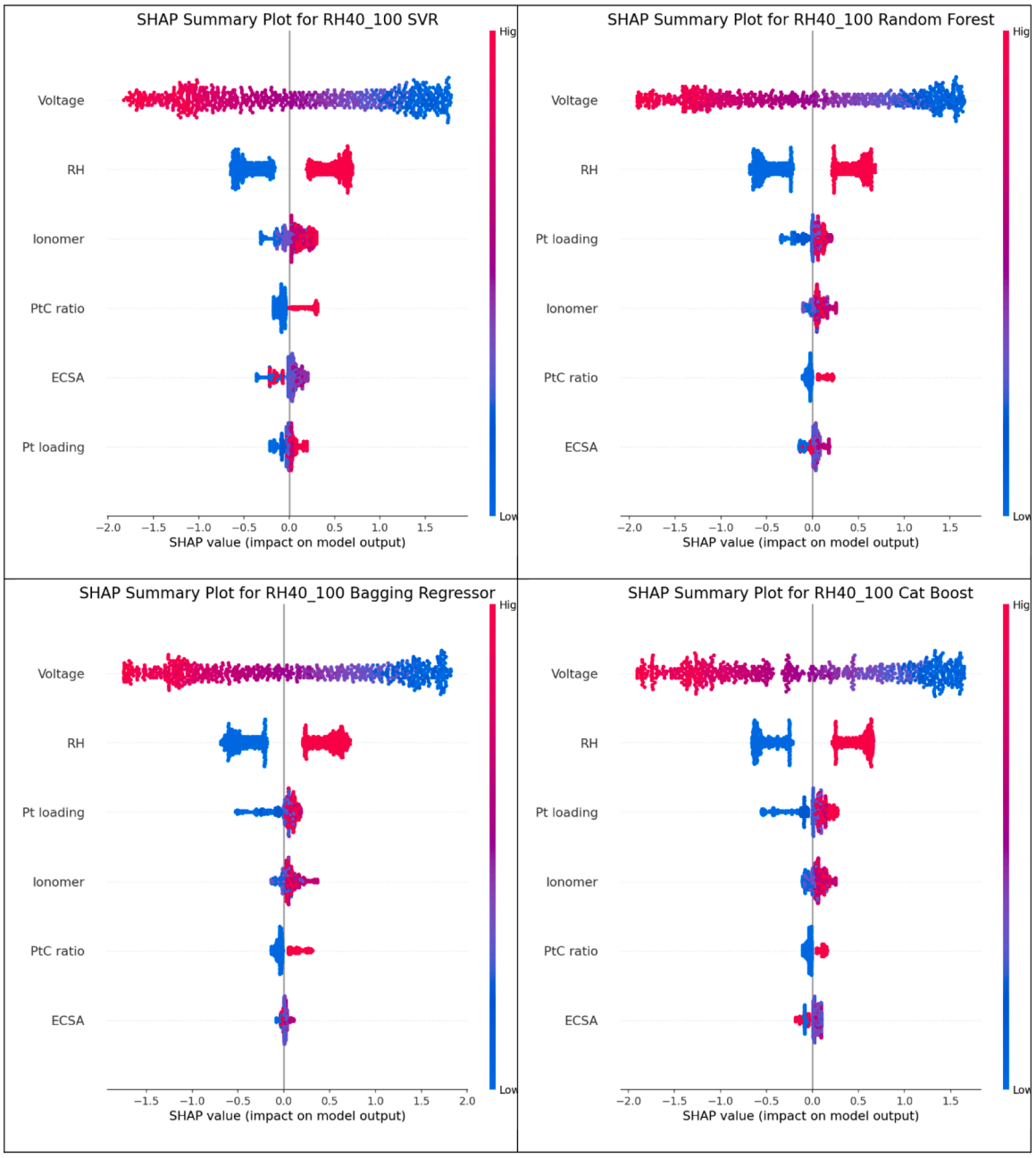


Fig. 5. Beeswarm SHAP plot for a) SVR, b) Random Forest, c) Bagging, and d) CatBoost, illustrating feature impacts with high values in red and low values in blue, across all features in the top-performing models.

current density prediction models. RH significantly affected proton transport resistance in the intermediate regime, while platinum loading and I:C mass ratio revealed optimal thresholds for enhancing efficiency and durability.

The models' decisions on using feature values for predicting current density were found to be realistic and consistent with findings reported in the literature. These interpretable ML models enable precise identification of key performance drivers, facilitating targeted improvements in PEFC design. Integrating fabrication data enhances the linkage between catalyst characterization and performance outcomes in PEFCs, revealing how precise manufacturing techniques impact fuel cell efficiency and durability.

Our approach, enhanced by explainable AI tools, is expected to be able to inform advanced fabrication processes, towards achieving optimal performance and extended operational life of fuel cells. Interpretable ML models, such as those employed, enable precise identification of key performance drivers, facilitating targeted improvements in PEFC design.

This insight-driven approach, implemented for PEFCs, should be adaptable to other applications, such as PEM water electrolysis. The models also lay the foundation for digital twins, enabling real-time monitoring and optimization of PEFC operation.

Future work could enhance this approach by integrating data-driven models with physics-based models, merging the computational efficiency of ML with the mechanistic accuracy of physical models to improve adaptability to dynamic conditions and long-term performance forecasting. Such a combined approach could enable more effective adaptation to changing operational conditions and improved long-term performance forecasting, ultimately maximizing fuel cell lifespan and efficiency. In addition, future work will include external validation

using independent datasets to assess the robustness and generalizability of the proposed framework.

CRediT authorship contribution statement

Ali Malek: Writing – original draft, Visualization, Validation, Methodology, Investigation, Data curation, Conceptualization. **Max Dreger:** Writing – review & editing, Methodology, Data curation. **Nima Shaigan:** Writing – review & editing, Methodology, Conceptualization. **Chaojie Song:** Writing – review & editing, Conceptualization. **Kourosh Malek:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Jasna Jankovic:** Writing – review & editing, Data curation. **Michael Eikerling:** Writing – review & editing, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work was supported by the German-NRC collaboration project.

Acknowledgment

The authors also greatly acknowledge researchers in IET-3 at the Forschungszentrum Jülich for valuable insights and contributions to this project. MD, KM, and ME also acknowledge the partial financial support from the European Union's Horizon Europe Research and Innovation programme, project DECODE under Grant Agreement No 101135537. AM acknowledges the grant for research exchange provided by Center for Advanced Simulation and analytics (CASA), Simulation and Data Science Lab for Energy Materials (SDL-EM) at the Forschungszentrum Jülich GmbH, taken place during Summer 2024. The authors would like to acknowledge Dr. Nada Zamel for sharing the experimental data and providing valuable feedback on this study.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.egyai.2025.100577.

Data availability

The original underlying data presented in the study are included in the article supplementary information further inquiries for access to data can be directed to the corresponding author.

References

[1] Olabi AG, Abdelkareem MA. Renewable energy and climate change. Renew Sustain Energy Rev 2022;158. https://doi.org/10.1016/j.rser.2022.112111.
[2] Rolnick D, Donti PL, Kaack LH, Kochanski K, Lacoste A, Sankaran K, Ross AS, Milojevic-Dupont N, Jaques N, Waldman-Brown A, Luccioni AS, Maharaj T, Sherwin ED, Mukkavilli SK, Kording KP, Gomes CP, Ng AY, Hassabis D, Platt JC, Creutzig F, Chayes J, Bengio Y. Tackling climate change with machine learning. ACM Comput Surv 2022;55:1–96. https://doi.org/10.1145/3485128.
[3] Wang Y, Pang Y, Xu H, Martinez A, Chen KS. PEM fuel cell and electrolysis cell technologies and hydrogen infrastructure development – a review. Energy Environ Sci 2022;15:2288–328. https://doi.org/10.1039/d2ee00790h.
[4] Tellez-Cruz MM, Escorihuela J, Solorza-Feria O, Compañ V. Proton exchange membrane fuel cells (PEMFCs): advances and challenges. Polymers-Basel 2021;13. ARTN 3064 10.3390.polym13183064.
[5] Zhang Y, Wang J, Yao Z. Recent development of fuel cell core components and key materials: a review. Energies 2023;16. https://doi.org/10.3390/en16052099.
[6] Soomro IA, Memon FH, Mughal W, Khan MA, Ali W, Liu Y, Choi KH, Thebo KH. Influence of operating and electrochemical parameters on PEMFC performance: a simulation study. Membranes 2023;13:259.
[7] Tawalbeh M, Alarab S, Al-Othman A, Javed RMN. The operating parameters, structural composition, and fuel sustainability aspects of PEM fuel cells: a mini review. Fuels 2022;3:449–74.
[8] Eslamibidgoli MJ, Huang J, Kadyk T, Malek A, Eikerling M. How theory and simulation can drive fuel cell electrocatalysis. Nano Energy 2016;29:334–61.
[9] Li Y, Zhou Z, Liu X, Wu W-T. Modeling of PEM fuel cell with thin MEA under low humidity operating condition. Appl Energy 2019;242:1513–27.
[10] Kulikovsky AA. Analytical modelling of fuel cells. Elsevier; 2019.
[11] Eikerling M, Kulikovsky A. Polymer electrolyte fuel cells: physical principles of materials and operation. Crc Press; 2014.
[12] Squadrito G, Maggio G, Passalacqua E, Lufrano F, Patti A. An empirical equation for polymer electrolyte fuel cell (PEFC) behaviour. J Appl Electrochem 1999;29:1449–55.
[13] Kim J, Lee SM, Srinivasan S, Chamberlin CE. Modeling of proton exchange membrane fuel cell performance with an empirical equation. J Electrochem Soc 1995;142:2670.
[14] Hao D, Shen J, Hou Y, Zhou Y, Wang H. An improved empirical fuel cell polarization curve model based on review analysis. Int J Chem Eng 2016;2016:1–10. https://doi.org/10.1155/2016/4109204.
[15] Pisani L, Murgia G, Valentini M, D'Aguanno B. A new semi-empirical approach to performance curves of polymer electrolyte fuel cells. J Power Sources 2002;108:192–203.
[16] Shen J, Du C, Yan F, Chen B, Tu Z. Two parameters identification for polarization curve fitting of PEMFC based on genetic algorithm. Int J Energy Res 2022;46:9621–33. https://doi.org/10.1002/er.7831.
[17] Thosar AU, Agarwal H, Govarthan S, Lele AK. Comprehensive analytical model for polarization curve of a PEM fuel cell and experimental validation. Chem Eng Sci 2019;206:96–117. https://doi.org/10.1016/j.ces.2019.05.022.
[18] Haji S. Analytical modeling of PEM fuel cell i–V curve. Renew Energy 2011;36:451–8.
[19] Kulikovsky AA. A physically–Based analytical polarization curve of a PEM fuel cell. J Electrochem Soc 2014;161:F263–70. https://doi.org/10.1149/2.028403jes.
[20] Muzaffar T, Kadyk T, Eikerling M. Tipping water balance and the Pt loading effect in polymer electrolyte fuel cells: a model-based analysis. Sustain Energy Fuels 2018;2:1189–96.
[21] Zhao D, Dou M, Zhou D, Gao F. Study of the modeling parameter effects on the polarization characteristics of the PEM fuel cell. Int J Hydrogen Energy 2016;41:22316–27. https://doi.org/10.1016/j.ijhydene.2016.09.112.
[22] Eikerling M, Kornyshev A. Modelling the performance of the cathode catalyst layer of polymer electrolyte fuel cells. J Electroanalyt Chem 1998;453:89–106.
[23] Zhang G-X, Song Y, Zhao W, An H, Wang J. Machine learning-facilitated multiscale imaging for energy materials. Cell Reports Phys Sci 2022;3. https://doi.org/10.1016/j.xcrp.2022.101008.
[24] Rehman A, Tariq S, Farrakh A, Ahmad M, Javeid MS. A systematic review of machine learning and artificial intelligence methods to tackle climate change impacts. In: 2023 International Conference on Business Analytics for Technology and Security (ICBATS); 2023. p. 1–7.
[25] Ding R, Chen Y, Rui Z, Hua K, Wu Y, Li X, Duan X, Li J, Wang X, Liu J. Machine learning utilized for the development of proton exchange membrane electrolyzers. J Power Sources 2023;556. https://doi.org/10.1016/j.jpowsour.2022.232389.
[26] Wang Y, Seo B, Wang B, Zamel N, Jiao K, Adroher XC. Fundamentals, materials, and machine learning of polymer electrolyte membrane fuel cell technology. Energy and AI 2020;1. https://doi.org/10.1016/j.egyai.2020.100014.
[27] Yu Y, Yu Q, Luo R, Chen S, Yang J, Yan F. Degradation and polarization curve prediction of proton exchange membrane fuel cells: an interpretable model perspective. Appl Energy 2024;365. https://doi.org/10.1016/j.apenergy.2024.123289.
[28] Liu L, Liu T, Ding F, Zhang H, Zheng J, Li Y. Exploration of the polarization curve for proton-exchange membrane fuel cells. ACS Appl Mater Interfaces 2021;13:58838–47. https://doi.org/10.1021/acsami.1c20289.
[29] Ahmed R, Shehab SA, Elzeki OM, Darwish A, Hassanein AE. An explainable AI for green hydrogen production: a deep learning regression model. Int J Hydrogen Energy 2024;83:1226–42. https://doi.org/10.1016/j.ijhydene.2024.08.064.
[30] Ko T, Kim D, Park J, Lee SH. Physics-informed neural network for long-term prognostics of proton exchange membrane fuel cells. Appl Energy 2025;382. https://doi.org/10.1016/j.apenergy.2025.125318.
[31] Batool M, Sanumi O, Jankovic J. Application of artificial intelligence in the materials science, with a special focus on fuel cells and electrolyzers. Energy and AI 2024;18. https://doi.org/10.1016/j.egyai.2024.100424.
[32] Kim J, Baek J, Choi M. Machine-learning-driven feature importance analysis for guiding the protonic ceramic fuel cell manufacturing. Array 2025;26. https://doi.org/10.1016/j.array.2025.100407.
[33] Wang B, Xie B, Xuan J, Jiao K. AI-based optimization of PEM fuel cell catalyst layers for maximum power density via data-driven surrogate modeling. Energy Convers Manag 2020;205. https://doi.org/10.1016/j.enconman.2019.112460.
[34] Rui Z, Ding R, Hua K, Duan X, Li X, Wu Y, Wang X, Ouyang C, Li J, Li T, Liu J. Design of proton exchange membranes with high durability for fuel cells: from the perspective of machine learning. J Memb Sci 2023;683. https://doi.org/10.1016/j.memsci.2023.121831.
[35] Ding R, Ding Y, Zhang H, Wang R, Xu Z, Liu Y, Yin W, Wang J, Li J, Liu J. Applying machine learning to boost the development of high-performance membrane electrode assembly for proton exchange membrane fuel cells. J Mater Chem A 2021;9:6841–50.
[36] Zhao J, Li X, Shum C, McPhee J. A review of physics-based and data-driven models for real-time control of polymer electrolyte membrane fuel cells. Energy and AI 2021;6:100114.

[37] Oviedo F, Ferres JL, Buonassisi T, Butler KT. Interpretable and explainable machine learning for materials science and chemistry. Acc Mater Res 2022;3:597–607.
[38] Bhatt U, Xiang A, Sharma S, Weller A, Taly A, Jia Y, Ghosh J, Puri R, Moura JM, Eckersley P. Explainable machine learning in deployment. In: Proceedings of the 2020 conference on fairness, accountability, and transparency; 2020. p. 648–57.
[39] Casalicchio G, Molnar C, Bischl B. Visualizing the feature importance for black box models, machine learning and knowledge discovery in databases: european conference, ecml pkdd 2018. Dublin, Ireland: Springer; 2019. p. 655–70. September 10–14, 2018, Proceedings, Part I 18.
[40] Albini E, Long J, Dervovic D, Magazzeni D. Counterfactual Shapley additive explanations, 2022 ACM Conference on Fairness. Account Transp 2022:1054–70.
[41] Alink R, Singh R, Schneider P, Christmann K, Schall J, Keding R, Zamel N. Full parametric study of the influence of ionomer content, catalyst loading and catalyst type on oxygen and ion transport in PEM fuel cell catalyst layers. Molecules 2020; 25:1523.
[42] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–97.
[43] Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016. p. 785–94.
[44] Ho TK. Random decision forests, proceedings of 3rd international conference on document analysis and recognition. IEEE; 1995. p. 278–82.
[45] Vovk V. Kernel ridge regression. Empirical inference: festschrift in honor of vladimir n. vapnik. Springer; 2013. p. 105–16.
[46] Ranstam J, Cook JA. LASSO regression. J British Surg 2018;105. 1348-1348.
[47] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y. Lightgbm: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 2017;30.
[48] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. Adv Neural Inf Process Syst 2018;31.
[49] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Comput 2006;18:1527–54. 10.1162/neco.2006.18.7.1527.
[50] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 2005;30:79–82.
[51] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30.
[52] Nembrini S, König IR, Wright MN. The revival of the Gini importance? Bioinformatics 2018;34:3711–8.
[53] Tucker AW, Luce RD. Contributions to the theory of games. Princeton University Press; 1959.
[54] mustermann_2021_svgdigitizer_1.pdf.
[55] Cui J, Hang H, Wang Y, Lin Z. GBHT: gradient boosting histogram transform for density estimation. In: International Conference on Machine Learning, PMLR; 2021. p. 2233–43.
[56] Mazzeo F, Di Napoli L, Carello M. Assessing open circuit voltage losses in PEMFCs: a new methodological approach. Energies 2024;17:2785.
[57] Niroumand AM, Homayouni H, Goransson G, Olfert M, Eikerling M. In-situ diagnostic tools for hydrogen transfer leak characterization in PEM fuel cell stacks part III: manufacturing applications. J Power Sources 2020;448:227359.
[58] Olbrich W, Kadyk T, Sauter U, Eikerling M, Gostick J. Structure and conductivity of ionomer in PEM fuel cell catalyst layers: a model-based analysis. Sci Rep 2023;13: 14127.