Applied Energy 386 (2025) 125529

Contents lists available at ScienceDirect

Applied Energy

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

Data-driven surrogate modeling for performance prediction and sensitivity analysis of transport properties in proton exchange membrane water electrolyzers

K. Ashoke Raman [a],*, Linus Hammacher [a], Hans Kungl [a], André Karl [a], Eva Jodat [a], Rüdiger-A. Eichel [a,b], Violeta Karyofylli [a]

[a] Institute of Energy Technologies , Fundamental Electrochemistry (IET-1), Forschungszentrum, Jülich, 52425, Germany
[b] Institute of Physical Chemistry, RWTH Aachen University, Aachen, 25062, Germany

ARTICLE INFO

Keywords:
Machine learning
Sensitivity analysis
PEM water electrolyzer
Numerical simulation

ABSTRACT

Proton exchange membrane electrolytic cells (PEMEC) are complex multivariate electrochemical systems that have emerged as a prominent technology for generating green hydrogen. To reduce costs and accelerate the commercial deployment of PEMEC, it is crucial to develop accurate predictive models that enable to capture the inherent nonlinearities of PEM electrolyzers efficiently. Therefore, in this study, we develop data-based surrogate models for PEMEC with catalyst layers having high (supported) and low (unsupported) electronic conductivity using support vector regression, extreme gradient boosting and artificial neural networks machine learning techniques focusing on the system's transport properties. These models are developed by using the datasets obtained from an analytical model and a physics-based one-dimensional numerical model of PEMEC. The dataset obtained from the one-dimensional model was split into datasets for supported and unsupported catalyst layers, based on the electronic conductivity of the anode catalyst. The performance prediction of these three models is evaluated and compared with physics-based modeling results. We find that both artificial neural network (ANN) and extreme gradient boosting (XGB) models perform well in predicting the cell current density. Therefore, the artificial neural network (ANN) model is selected to perform parametric analysis to investigate the effect of operating conditions and transport properties of the anode side. Both shapely additive explanations (SHAP) and sensitivity analysis reveal that the operating temperature is the most important parameter affecting the performance of the proton exchange membrane electrolytic cell. For supported catalyst layers, the influence of membrane thickness is greater than the catalyst's electronic conductivity. However, in the case of unsupported catalysts layers, the SHAP values for electronic conductivity are found to be larger than membrane thickness.

1. Introduction

In pursuit of a transition towards a sustainable energy economy, the production and efficient utilization of green hydrogen is central to shaping the energy landscape [1]. To this end, PEMECs have emerged as a promising technology owing to their high efficiency, modular design, lean balance-of-plant layout due to lower gas crossover, and high purity of produced hydrogen [2]. PEMECs utilize a proton-conducting membrane as the electrolyte which helps in rapid proton transport between anode and cathode, thereby achieving high current densities. These features can make PEMECs a crucial component in the electrical grid as they offer fast response times. When compared with alkaline electrolyzers, PEMECs offer higher gas purity [3] and current densities [4].

In order to utilize PEMECs effectively in real-time situations, it is pertinent to be able to predict the system behavior under various operating conditions. PEMECs operate through a complex interplay between electrochemical reactions, mass transport, heat transfer, and fluid flow across different spatial and time scales [5]. Physics-based modeling of PEMECs helps in revealing insights into the workings of these complex charge and transport phenomena. This is accomplished by solving a set of governing differential equations that describe the physical processes occurring within the electrolyzer. However, these calculations can be computationally intensive for detailed modeling of large-scale systems in real-time conditions. To address this potential drawback, data-based modeling has emerged as an effective cost-reducing alternative. These

---

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

https://doi.org/10.1016/j.apenergy.2025.125529
Received 23 September 2024; Received in revised form 5 February 2025; Accepted 11 February 2025

Nomenclature

AL	Active Learning
ANN	Artificial Neural Network
BPNN	Back Propagation Neural Networks
CFD	Computational Fluid Dynamics
CGCNN	Crystal Graph Convolutional Neural Network
CL	Catalyst Layer
DFT	Density Functional Theory
ERT	Extremely Randomized Trees
GDL	Gas Diffusion Layer
HER	Hydrogen Evolution Reaction
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 Menbrane
PEMEC	Proton Exchange Membrane Electrolytic Cell
PEMFC	Proton Exchange Membrane Fuel Cell
R^2	Coefficient of determination
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error
SA	Sensitivity Analysis
SHAP	Shapely Additive Explanations
SI	Sensitivity Index
SR	Split Ratio
SVM	Support Vector Machine
SVR	Support Vector Regression
XGB	Extreme Gradient Boosting

models rely on large datasets obtained from experiments or simulations and learn the relationship between input parameters, such as operating conditions, material properties and design parameters, and output targets. Using statistical techniques they establish a relationship between input parameters and output targets which enables these models to make predictions, perform system optimization, and identify unknown patterns. In particular, these models have proven to be useful for online motoring [6], real-time fault detection, and prognostics. Owing to their empirical nature, these models can be directly integrated with control systems to regulate the operating conditions dynamically based on the real-time output of the electrolyzer and ensure optimal performance.

A major drawback of the current state of PEMECs is the usage of noble metals as catalysts for the hydrogen evolution reaction (HER) and oxygen evolution reaction (OER). Extremely randomized trees (ERT) were used to screen 26 key features such as atomic radius, electronegativity, lattice constant, etc., to develop a surrogate model based on back propagation neural networks (BPNN) from a database generated from DFT calculations. By combining machine learning potential (MLP) and crystal graph convolution neural network (CGCNN), a universal ML framework was developed to screen 43 alloys out of 2974 potential candidates showing ideal HER activity [7]. Unlike the HER, the OER for PEMECs is a 4-electron transfer reaction which makes it the rate-determining step for the electrolyzer. Expensive Iridium-based oxides are typically used as the anode catalyst. ML techniques have been used to screen these systems to obtain configuration which leads to optimized activity. A global geometry surface optimization was performed using the Gaussian approximation potential (GAP) methods to obtain the best surface configurations for rutile IrO_2 [8]. Similarly, active learning (AL) has been deployed in conjunction with DFT calculations to synthesize 75 candidates of IrO_3 polymorphs from 956 unique structures [9].

While electrocatalysts are critical for the functioning of PEMECs, their performance is sensitive to varying physical and chemical processes occurring at widely different length scales. The reaction zone in the electrode is nano/micro scale [10,11], gas diffusion through PTL at micro/millimeter scales, and the entire system, at both the single cell and stack level, has dimensions in meter scales. Parameters ranging from catalyst properties, membrane-electrode assembly (MEA) dimensions, and operating conditions have been shown to influence the performance of PEMECs [5,12]. Hence, it is effective to use ML techniques which consider parameters across different scales to predict and optimize electrolyzer performance. Weka ML [13] is used to generate a training database by modifying X-ray computed tomography images of the anode GDL. This ML technique computed the oxygen content of the entire diffusion layers directly through image data. Decision trees and random forest algorithms have been deployed to optimize the MEA. The models trained over a database generated from previously published articles have prescribed optimal catalyst loading, ionomer/carbon ratios, support materials for anode and cathode, and appropriate pore structure of GDL [14,15]. Similarly, polynomial regression [16] has been employed to furnish optimal cell design parameters based on hydrogen production rate, catalyst area, and type of cell design. At the stack level, neural networks have shown potential in being able to predict the stack efficiency and hydrogen flow rate for 10 cells connected in series [17]. More recently, Chen et al. [18] proposed a novel frame work using knowledge-integrated machine learning to advance the development of PEMEC.

From the above discussions, it is evident that ML has been predominantly applied at an atomistic scale to screen electrocatalysts for PEMECs. This is accomplished principally from DFT-generated databases. However, the need to optimize individual components such as MEAs, GDLs, and flow channels is also crucial for the efficient operation of a complex multivariate system like PEMEC. This can be achieved efficiently through the black-box nature of data-driven models. This black-box attribute of ML techniques facilitates in coupling multiple parameters, which govern different individual components (on multiple scales), to predict performance characteristics. Owing to the lack of such studies in the literature [19] for PEMECs, there is a need to employ physics-based models or experiments to generate databases and expedite the use of ML for optimizing individual components or the entire cell. Therefore, this paper aims to employ machine learning to model the performance of PEMEC on a cell level using data generated from macroscale physics-based models [12,20]. Thereby, shifting the focus from the existing body of work on ML-based screening of electrocatalysts for PEMEC. This is unlike PEMFCs [21], where ML has been utilized to a great extent on both individual components, as well as the entire PEMFC system/stack for system optimization and cost reduction. The PEMEC is governed by a large set of parameters including operating conditions, material, transport, and electrochemical properties. Given the critical importance of the anode side wherein the OER takes place, we focus our attention in this initial work on the operating conditions and transport properties of the anode side. In this context, we separately investigate the role of supported and unsupported anode catalyst layers on cell performance. Supported catalyst layers use a support material on which the catalyst particles are dispersed. In these systems, catalytic activity and electronic transports is predominantly undertaken by the catalyst particles and support material, respectively. This approach results in minimal usage of expensive noble materials and enhanced conductivity [22]. However, most of the support materials are unstable under anodic conditions for long term operations. Unsupported catalyst layers circumvent this problem, albeit at the cost of expensive catalyst material. The multi-functionality of unsupported catalyst layers is attributed to its ability to perform both the catalytic activity and electronic transport. For these type of catalyst layers, the electronic conductivity is influenced by the amount of ionomer content. Increase in ionomer content may lead to the reduction in the number of contact points between the solid particles, thereby, increasing the electronic resistance. We attempt to capture the interaction on how these two types of catalyst layers respond to varying operating conditions and anode transport properties. As electronic conductivity is one of the parameters which depend on the type of the catalyst layers, we divide the anode catalyst layer as supported and unsupported [23] based on the values of the anode electronic conductivity. To the best of our knowledge, there is no previous study applying different ML techniques and comparing them in the context of an entire PEMEC model from data-generated from physics-based models.

The paper is organized as follows: We begin with a brief description of the three different ML techniques 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. Method

Three different types machine learning algorithms, each with a unique mathematical foundation, are selected in this study to investigate their comparative value in making performance prediction for PEMEC. The support vector regression (SVR) is a classical statistical technique which utilizes convex optimization by transforming the input data into a higher dimensional feature space using kernal functions [24]. On the other hand, the constituting statistical component of extreme gradient boosting (XGB) is the decision tree, which is aggregated sequentially, to built an accurate ensemble model. The accuracy of this model is enforced by minimizing the errors using gradient boosting. Artificial neural networks (ANN) rely on a set of layers which are interconnected by biologically inspired non-linear activation functions called *neurons*. We aim to investigate on how these three fundamentally different algorithms perform in predicting the performance of a highly non-linear PEMEC system. In this section, we briefly introduce the prominent features of the three ML methods used in this study. We then proceed with describing data acquisition and preprocessing.

2.1. Support vector regression (SVR)

Support vector regression [25] is a machine learning technique utilized in predicting continuous data. This method borrows concepts from the Support Vector Machines (SVM) method, wherein a hyperplane in high dimensional feature space is used to classify two distinct classes. Similar to SVM, SVR utilizes *kernel* functions to enlarge the input feature space which enables the creation of a hyperplane in the resulting multidimensional space. Thereby, the non-linearities arising from the relationship between initial input parameters and output target are linearized in higher dimensions. Common *kernel* functions include linear, polynomial, Gaussian radial basis function (RBF), and sigmoid. The hyperplane is enclosed within a decision boundary, whose extent is controlled by the hyperparameter ϵ. The objective is to tune the SVR model such that it encloses the maximum number of training data points within the decision boundary around the predicted value. The data points lying out of the decision boundary contribute to the model's error. Apart from ϵ and the choice of the *kernel* function, SVR has another hyperparameter C which controls the bias–variance trade-off of the model.

2.2. XGBoost (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. Mathematically, it optimizes the following objective function for a given dataset $(x_i, y_i)_{i=1}^{n}$:

$$Obj = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{1}$$

where the loss function $\mathrm{L}(y_i, \hat{y}_i)$ is a squared loss function expressed by

$$L(y_i, \hat{y}_i) = (y_i - \hat{y}_1)^2 \tag{2}$$

and $\Omega(f_k)$ is the regularization term for the kth tree. The final prediction by the model is made by an additive combination of the output of all the trees

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \tag{3}$$

2.3. Artificial neural networks (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. Fig. 1 shows a representative architecture for ANNs. 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 as shown from the interconnections in Fig. 1. Mathematically, the output signal of a single neuron can be expressed as follows:

$$z_i^l = \sum_{r=1}^{m} (w_{ir}^l x_r^{l-1} + b_i^l) \tag{4}$$

$$a_i^l = \phi(z_i^l) \tag{5}$$

Here, the indices l and i represent layer number and corresponding neuron index, respectively. The output signal (z_i^l) is the input obtained by a linear combination of weights (w_{ir}^l: w_{i1}^l, w_{i2}^l, w_{i3}^l, … , w_{im}^l) and corresponding m activation values from the previous layer (a_i^{i-1}). The term b_i^l is the bias which controls the activity of the corresponding node. Activation values are computed by applying the activation function ϕ on z_i^l. The most commonly used activation functions are the sigmoid, rectified linear unit (ReLU), tan hyperbolic and softmax functions. These non-linear activation functions allow the neural networks to learn the non-linearities present in a given database. For the first hidden layer, z_i^l corresponds to the values of the input layer. This transmission of activation values from the input layer to the output layer is known as *forward propagation*. The prediction obtained from the output layer

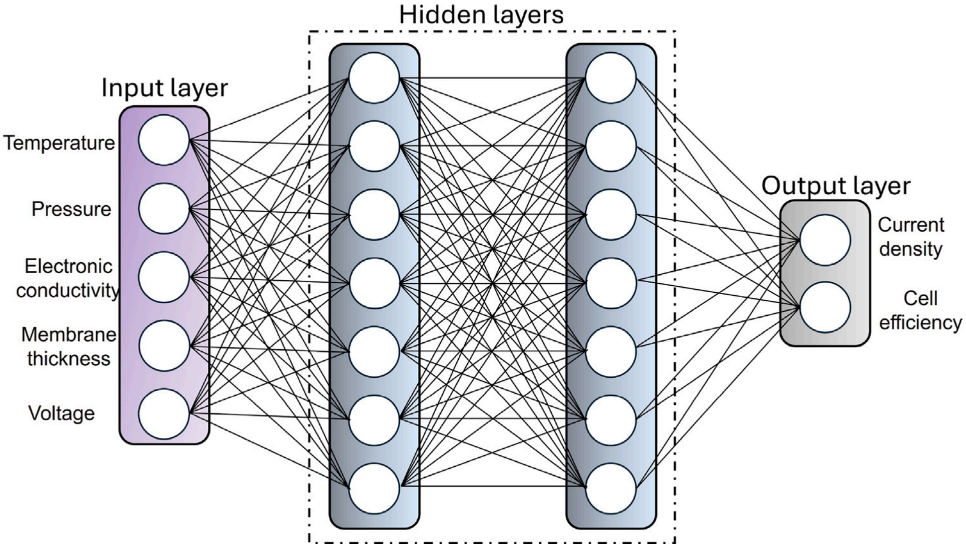

Fig. 1. Schematic representation of architecture for an artificial neural network. The architecture of a typically consists of an input layer, two or more hidden layers, and an output layer.

($\hat{y}$) are compared with the desired output (y) for the input data point. For a dataset with n sample points, the cost function is given as:

$$J(w,b) = \frac{1}{n}\sum_{i=1}^{n} \mathscr{L}(\hat{y}, y) \tag{6}$$

where $\mathscr{L}$ is the loss function which computes the error for a given data point. The parameters w^l_{ir} and b^l_i are updated by optimizing the cost function through the gradient decent. The weights and biases of each neuron are adjusted such that the associated error is minimized in each iteration. This correction of weights and biases has a general form expressed as:

$$W^l = W^l - \alpha dW^l \tag{7}$$

$$B^l = B^l - \alpha dB^l \tag{8}$$

where W^l is the weight matrix comprising weights for all the nodes inside the layer l, and B^l is the corresponding bias vector. The term α is a positive constant known as the learning rate. This correction process flows in the opposite direction, starting from the output layer, and is known as *backpropagation* [26]. The weights, biases, number of hidden layers, neurons per layer, and the learning rate constitute the hyperparameters of ANNs.

2.4. Data acquisition and preprocessing

At the core of any machine-learning model is the quality of data on which the model is developed. As such, it is crucial to gather reliable and clean data to circumvent inaccurate predictions and generalize well on unseen data. Usually, data for developing data-based models is gathered directly from actual PEMEC experiments [17,27] or through simulating physics-based models by varying a range of parameters [28,29]. In this study, we develop data-based models from data gathered from two different physics-based models. These models have been validated against existing experimental literature. Dataset 1 is generated by simulating a zero-dimensional analytical model [20] for PEMEC. We have varied the applied cathodic pressure, anodic pressure, exchange current density, the separator thickness, and applied current density to calculate the output cell voltage. The range of these input parameters is shown in Table 1. The second dataset is generated from a one-dimensional two-phase, non-isothermal CFD model of PEMEC [12]. This model accounts for the description of mass, charge, species, and heat transport. Current density data obtained from this model encompasses the complex non-linear coupling between the set of nine conservation equations [12]. The 1D model, which was developed by Salaberri [12], has been validated against the experimental results of Chandesris et al. [30] and Debe et al. [31] for different physicochemical and geometric parameters of the membrane electrode assembly. The polarization curves obtained from the model showed good agreement with the experimental data. Therefore, the synthetic data generated from the physics based model serves as an experimentally validated database to develop accurate data-driven surrogate models. The input parameters and their range of variation are presented in Table 2. The input features comprise parameters influencing the operating conditions (T^{in} and P^a) and design parameters influencing effective transport properties (σ^a_{cl}, δ^a_{cl}, δ^a_{ptl} and δ_{mem}). The operating parameters (T^{in} and P^a) also effect several transport properties such as the molecular diffusivity of water vapor, effective protonic conductivity, effecttive water diffusivity, liquid saturation and electrochemical reaction rates. A central parameter in which the investigated experimental data sets for supported [30] and unsupported catalysts differ is the electrical conductivity. Reported values for the effective electronic conductivity of anodic catalyst layers in PEMECs show a variation across several orders of magnitude between supported and unsupported catalyst layers. For example, Mandal et al. [23] found that the through-plane electronic conductivity of the unsupported IrO_x catalyst layer is in the range of 10^{-3}-10^{-5} S/cm for different ionomer fractions and relative humidity values [23]. Therefore, based on the above mentioned literature, in this study we have selected the electronic conductivity as the transport property to split the second dataset into two parts to account for supported and unsupported cases. However, it should be noted that other important factors such as the active catalyst surface area, catalyst loading and ionomer content also influence the overall electronic conductivity of the catalyst layers. The values of the considered δ_{mem} correspond to the thickness of Nafion 212, Nafion 115, and Nafion 117. There are two output variables in these datasets: current density and efficiency. The efficiency (η_e) is defined as the ratio between the energy required to break one mole of water under equilibrium conditions to the energy required to decompose the same in non-equilibrium conditions. It is given by the following expression [32] :

$$\eta_e = \frac{2FE_{tn}}{2F(E_{tn} + V_{cell} - E^{rev})} \tag{9}$$

where E^{rev} is the reversible cell voltage, V_{cell} is the applied cell voltage and E_{tn} is the temperature-dependent thermoneutral voltage expressed as:

$$E_{tn} = 1.485 - 1.49 \times 10^{-4}(T - 273.15) - 9.84 \times 10^{-8}(T - 273.15) \tag{10}$$

Table 1
Range of input variables for the 0D model.

Inputs (Feature vectors)	Unit	Range of values
Applied cathodic pressure, P^c	bar	1.0, 4.0, 6.0
Applied anodic pressure, P^a	bar	1.0, 4.0, 6.0
Exchange current density, i_0	A cm^{-2}	1.0×10^{-9}, 2.602×10^{-10}, 1.0×10^{-12}
Separator thickness, d_{sep}	cm	0.0057, 0.0127, 0.021

Table 2
Range of input variables for the 1D numerical model.

Inputs (Feature vectors)	Unit	Range of values
Operating temperature, T^{in}	°C	40, 50, 60, 70, 80
Operating anodic pressure, P^a	bar	1.0, 2.0, 3.0
Anode catalyst layer thickness, δ^a_{cl}	μm	8, 12, 16, 20
Anode porous transport layer thickness, δ_{ptl}	μm	200,400, 600, 800
Membrane thickness, δ_{mem}	μm	50, 127, 180
Anode catalyst electronic conductivity (unsupported), σ^a_{cl}	S m^{-1}	0.0035, 0.0085, 0.035, 0.085
Anode catalyst electronic conductivity (supported), σ^a_{cl}	S m^{-1}	3.5, 35.0, 350

Since the data is directly generated from physics-based models, the input variables shown in Tables 1 and 2 are directly used as the input features for developing the data-based models.

Both, input features and output variables have widely different scales. Features with larger scales would lead to biasing of the machine learning model towards these input variables. This necessitates the need to scale these parameters before training the model. We have normalized the input features and output variables between 0 and 1. Normalization is employed for the SVR and ANN models, whereas we have omitted this step for XGB models [33]. Tree-based models are generally less sensitive to feature scaling as they are split based on the threshold feature values.

3. Results and discussion

We begin our discussions by first investigating the performance of developed data-driven models for dataset 1 in Section 3.1. Here we define three statistical evaluation metrics which are generally used to quantify the performance of machine learning models. After discussing data-driven models generated from an analytical model [20], we next explore the performance of machine-learning models developed from the data generated by a non-linear one-dimensional CFD model [12] in Section 3.2.

3.1. Dataset 1: Zero-dimensional analytical model

A dataset consisting of 1600 data points originating from the analytical model is randomly distributed into a training set and a test set. The parameters used to generate this dataset are given in Table 1. We have fixed the split ratio to 70:30. The training set is used to develop the machine learning models. During this process, the hyperparameters for each algorithm are tuned so that the resulting models provide accurate predictions for the 1200 data points of the training set. The mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination (R^2) are used to quantify the performance of the data-driven models. The mathematical expressions of MAE, MSE, and R^2 are given below:

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2} \tag{12}$$

$$R^2 = 1 - \frac{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_j)^2} \tag{13}$$

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

Fig. 2(a) illustrates the relationship between predicted voltage values and the ground truth for the training data set. We observe that all three models show excellent voltage predictions, with the ANN model showing the best RMSE values of 10^{-3}. The R^2 values of the SVR, XGB, and ANN models are measured to be 0.999 respectively. The closer the value of R^2 is to one, the stronger the corresponding correlation between the model predictions and actual values. As such, all the data-driven models show an excellent correlation between predicted and true voltage values.

The next step after building a model based on the training dataset is to evaluate its prediction capability on unseen data. This step is accomplished by using the developed models to make predictions on the test dataset, which in this case is 400 data points. Fig. 2(b) shows the absolute error between the predicted and actual voltage values for the different test samples. We notice the least performing model on the test dataset is the SVR model with a maximum absolute error of 0.037 V. It is clearly observed that several test samples have an absolute error greater than 0.01 for the SVR model. When compared to the SVR model, both XGB and ANN models produce better predictions on the test data. Furthermore, the maximum absolute error for XGB and ANN are calculated to be 0.0074 V and 0.0044 V, respectively. Therefore, the predictive performance of the ANN model is best on the training dataset.

To further elucidate the difference between the three methods, Fig. 2(c) illustrates a polarization curve with a randomly selected working condition: $P^c = 4.0$ bar, $P^a = 1.0$ bar, $d_{sep} = 0.0127$ cm, and $i_0 = 2.602 \times 10^{-10}$. In line with our observations in Fig. 2(b), we notice that the SVR model shows deviations in voltage prediction when compared to the analytical model as it is subjected to unseen data. The departure from the ground truth is particularly large as it over-predicts the voltage at lower current densities (0.0–0.5 A/cm^2). This difference reduces as the current density approaches to 1.0 A/cm^2, beyond which the model begins the under predict the ground truth. Both XGB and ANN show excellent voltage predictions when compared with the analytical model. The cell voltage values, as well as the trend of the polarization curves are captured well by XGB and ANN models. This showcases the superior learning ability of both these models to learn patterns and output values from the training dataset. The mean absolute error of XGB and ANN models computed from the test data is found to be 0.0016 V and 0.0010 V, respectively.

3.2. Dataset 2: One-dimensional CFD model

The machine learning models developed so far learn from the data generated from an semi empirical model. Although derived from physics-based governing equations, 0D models have reduced dimensionality and represent lower complexity of the system. Training ML models using the output of such models is relatively easier and requires

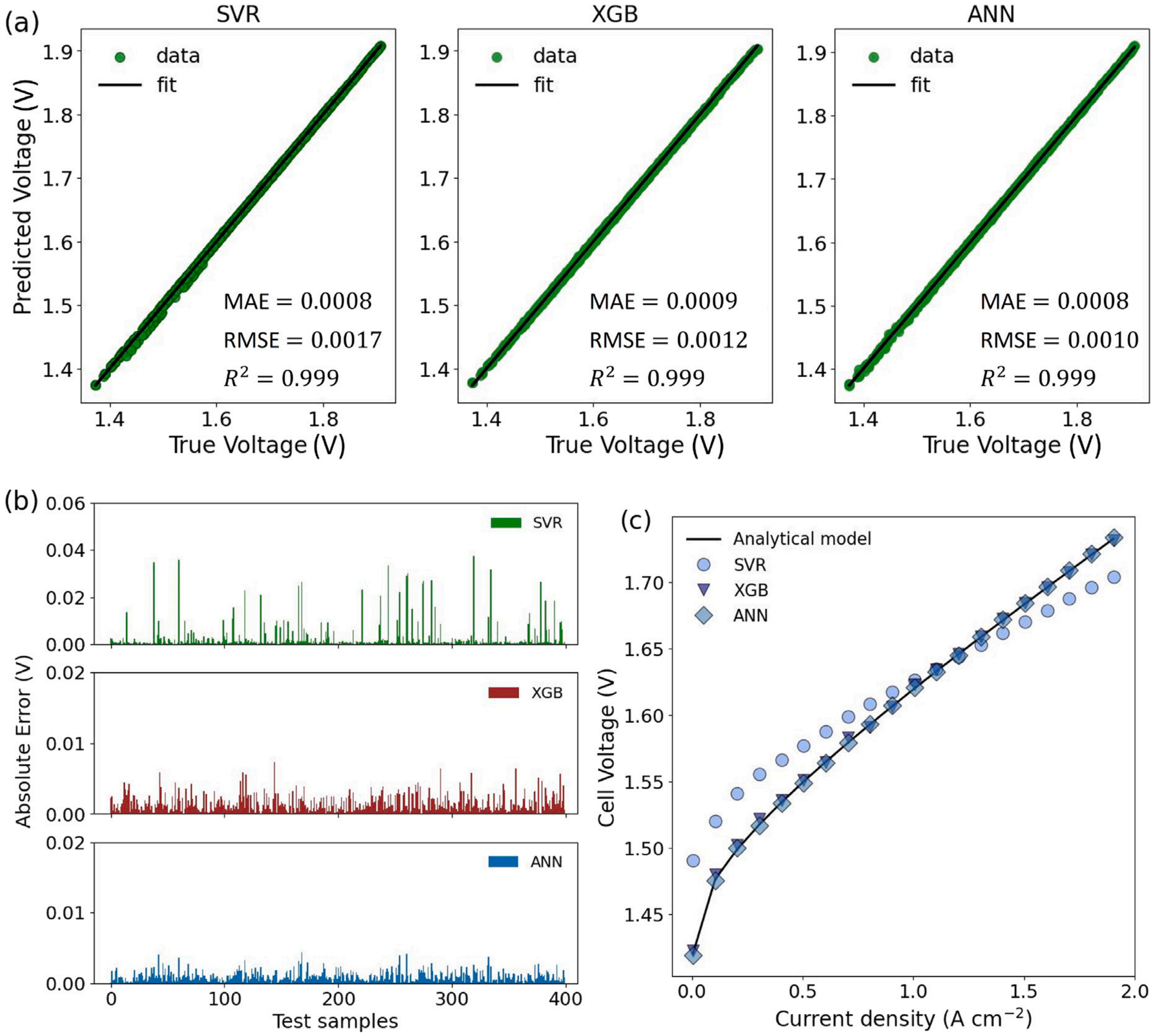

Fig. 2. (a) Relationship between voltage predicted from data-based and 0D analytical model values for the training data set (b) Residual between voltage predicted from data-based and 0D analytical model in the test set. (c) Comparison between the different data-based models with the analytical model for a test condition: $P^c = 4.0$ bar, $P^a = 1.0$ bar, $d_{sep} = 0.0127$ cm, and $i_0 = 2.602 \times 10^{-10}$ A cm^{-2}.

less data because the input–output relationship is less intricate. The 1D models are more complex, multivariate partial differential equations which contain large non-linear interactions. To develope data-driven models from such equations, we require detailed datasets with include a wide range of boundary conditions and parameter variations. As such, the performance capability of ML models to learn and predict non-linearities associated with a data set obtained from coupled set of multivariate non-linear partial differential equation has not yet been tested. In this section, we move from the relatively simple 0D model and focus on developing such data-based models and evaluating their predictive capacity by learning from non-linear datasets. This dataset is generated from a one-dimensional physics-based model composed of nine coupled conservation equations [12].

3.2.1. Data-driven model performance

Since Dataset 2 is a multi-output regression problem, we have developed separate models for predicting current density and efficiency using the SVR and XGB algorithms. However, a single model is developed for the ANN algorithm by building a network with two output nodes. As the dataset generated from the 1D-numerical model is segregated into supported and unsupported datasets, separate machine-learning models are developed for each case. The data set consists of 36 703 and 48 943 data points for the supported and unsupported cases, respectively. Further, each of these datasets has been split in the ratio of 80:20 for the training and testing datasets. The optimal hyperparameters for the SVR model for current density prediction are as follows: C = 15.0, $\epsilon = 0.004$, and $\gamma = 10$. The radial basis function is employed as its kernel. For predicting efficiency, the values of SVR model hyperparameters C and ϵ were set to be 30 and 0.01, respectively. Similarly, the optimal number of estimators was set to be 700 for the XGBoost model. The learning rate was set to be 0.015. The hyperparameters for ANN are the same for current density and efficiency predictions. A deep neural network architecture was built with three hidden layers consisting of 60, 45, and 25 units, respectively. The callback functionality was implemented to optimize the neural network model. The mean squared error criterion is used as the loss function with the optimal learning rate found to be 0.005. This assists in halting the training process once the training loss falls below a predefined loss value. We have set this value to be 10^{-6}. The hyperparameters of the employed models, such as the number of neurons, number of hidden layers, learning rate, ϵ, number of estimators, etc. were obtained by using the grid search hyperparameter tuning technique. This method evaluates the model's performance for every combination of a predefined set of hyperparameters and identifies the best combination. All the models were trained on a laptop with 16 GB RAM and equipped with an intel core 12th generation i5 processor. The training time for

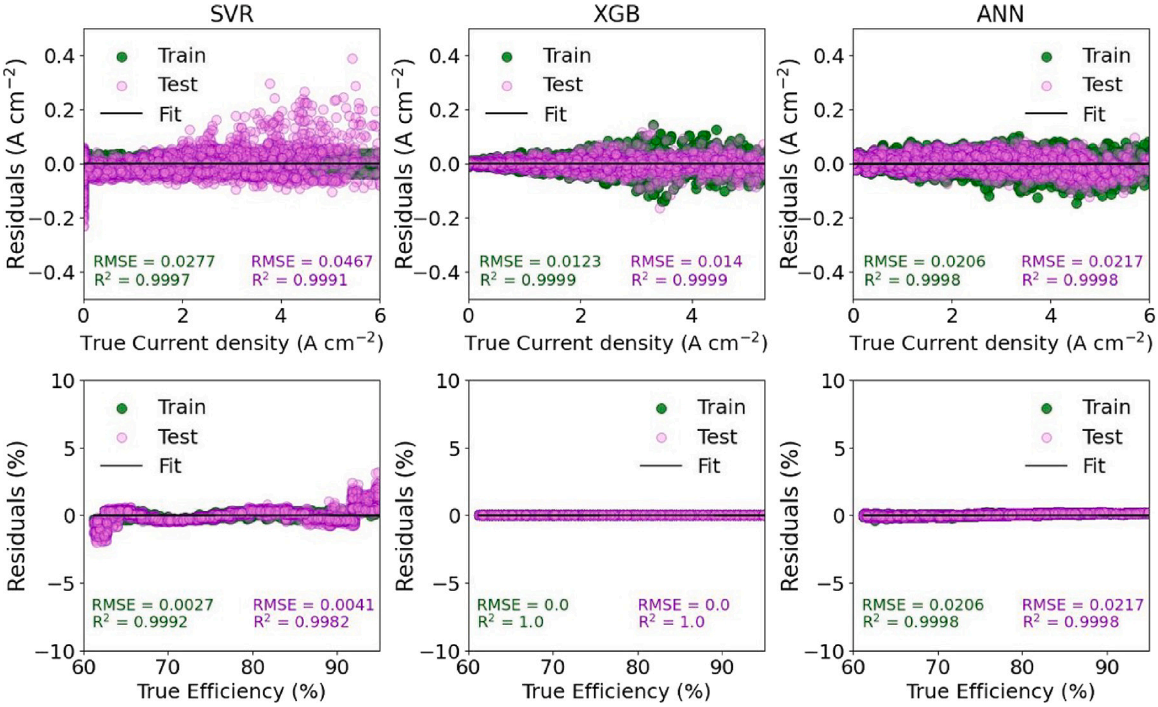

Fig. 3. Residual plots showing the comparison between the predicted current densities (top row) and efficiencies (bottom row) with those obtained from the 1D numerical model for the supported anode catalyst. The model's performance on the training and test set are shown in each plot by the RMSE and R^2 values.

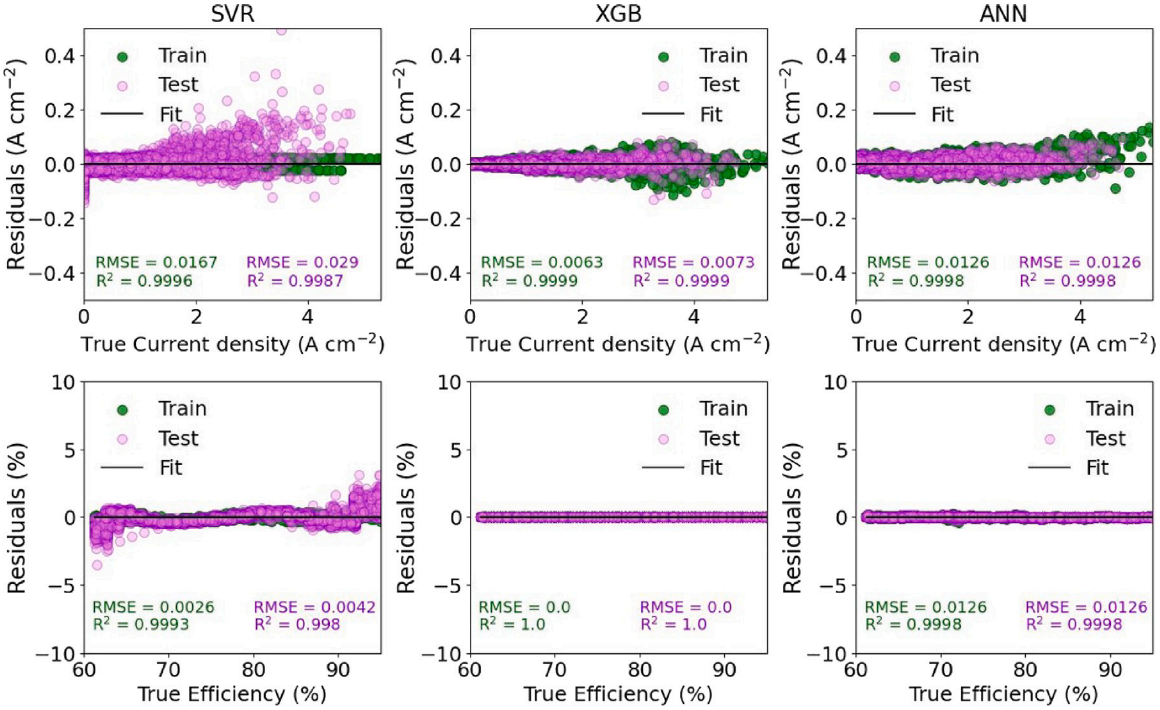

Fig. 4. Comparison between the predicted current densities (top row) and efficiencies (bottom row) with those obtained from the 1D numerical model for the unsupported anode catalyst. The model's performance on the training and test set are shown in each plot by the RMSE and R^2 values.

SVR, XGB and ANN models was approximately 1 min, 4 s and 23 min, respectively.

We first begin our discussion on model performance for supported catalysts. Fig. 3 illustrates the predicted outcomes of the current densities (top row) and efficiency (bottom row) from the SVR, XGB, and ANN machine learning algorithms. A residual is defined as the difference between the ground truth and model prediction. All three models appear to fit well on the training dataset with R^2 scores greater than 0.99. This establishes a good correlation between input features and the output variable. When compared with the SVR model, both XGB and ANN have shown better performance in predicting current density over the test data with the RMSE of 0.014 and 0.0217, respectively. Even though the SVR model has shown good performance on the training data, we notice a number of outliers for the test data. The value of

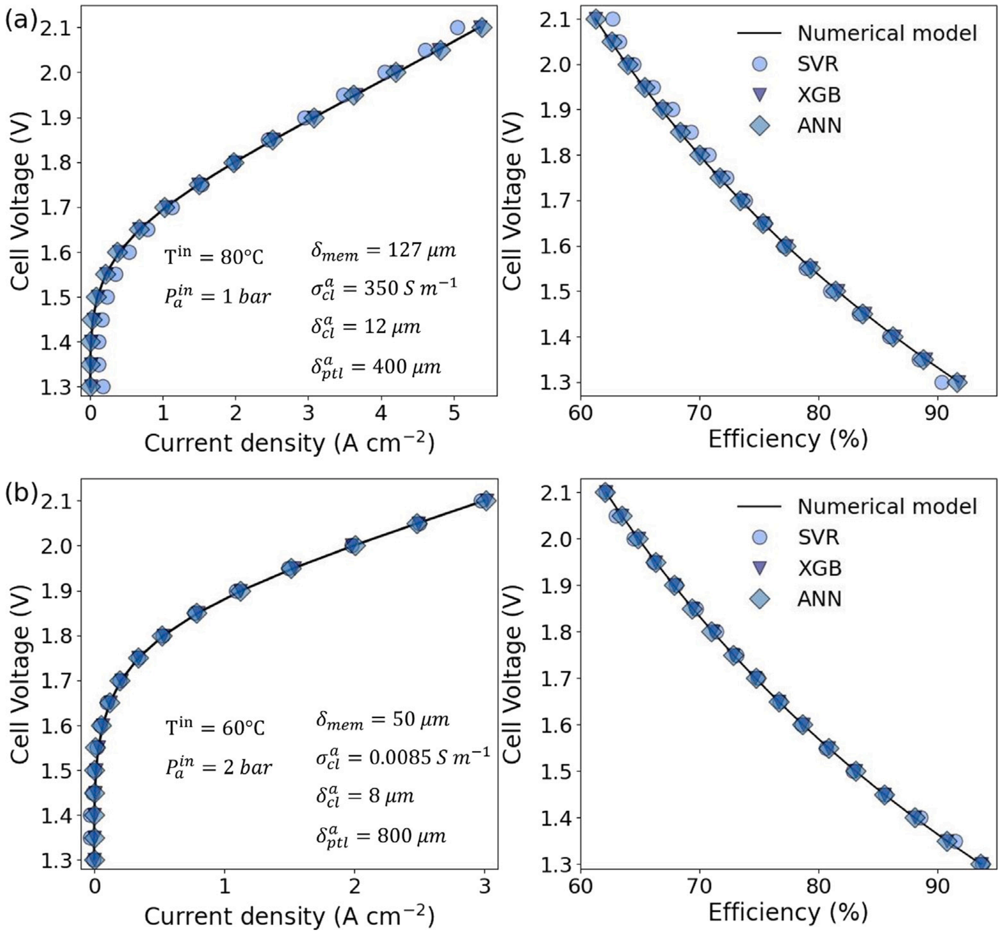

Fig. 5. Polarization curves and efficiencies of the three methods under two different conditions obtained from the data generated from the 1D model. The operating conditions are given in polarization curves.

the maximum residual on the test data is found to be 0.518 A/cm^{-2} when compared with 0.04 A/cm^{-2} on the training data. This shows that the SVR method has a good *bias* but suffers from high *variance*. The RMSE values for the training and test datasets are similar for both XGB and ANN models. Moreover, these models show a similar distribution of residuals for the training and test datasets around the fit line. Therefore, both these models exhibit low *bias* and *variance* in predicting current densities. We next assess the performance of data-based models in predicting efficiency, which is defined by Eq. (9). The XGB and ANN models show good performance with R^2 greater than 0.999 on the test data. The SVR model shows deviations from the fit line at lower efficiency (60.0–70.0%). At higher (>90.0%) efficiencies, we notice outliers in the test data for the SVR model.

The performance of the three data-driven models for the unsupported catalyst is shown in Fig. 4. When compared with the supported catalyst, the models show RMSE values for current density prediction reduced by nearly 50% on training and test datasets for unsupported catalyst layers. The maximum residual value for SVR, XGB, and ANN models on the test data was found to be 0.49 A/cm^{-2}, 0.086 A/cm^{-2}, and 0.088 A/cm^{-2}, respectively. This can be attributed to the fact that the range within which σ_{cl}^a is varied is between 0.0035–0.085 S/m for unsupported catalysts. The relative performance between the three models shows similar behavior as exhibited for supported catalysts. The SVR model continues to show high *variance* owing to a large number of outliers on the test data. Both XGB and ANN models continue to show better prediction performance with the XGB model performing slightly better (RMSE = 0.0073) than the ANN model (RMSE = 0.0126).

While RMSE quantifies the average magnitude of errors, it is also important to quantify how large the errors are relative to the actual values. This error metric is quantified by the mean absolute percentage error (MAPE), which is given by:

$$MAPE = \frac{1}{n}\sum_{j=1}^{n} \left| \frac{y_j - \hat{y}_j}{y_j} \right| \times 100 \tag{14}$$

MAPE is particularly important to indicate if the data-driven model consists of a large number of outliers. As this metric computes the relative error in percentage, the error arising from the outliers of the model contribute predominantly to the MAPE, leading to its overamplification. To further probe into the model's performance, Table 3 illustrates the MAPE values obtained from the test dataset for both supported and unsupported catalyst layer scenarios. The high values of MAPE for the SVR model in predicting the current density is indicative of the large number of the outliers seen from Figs. 3–4 (top row). The SVR model relies on the kernel function to transform the input features into a higher-dimensional space assuming that a linear relationship is established. When dealing with highly non-linear data, as obtained from the 1D model, the RBF is not capable of mapping the input features to establish a linear relationship with current density. As a result, we observe a large number of data points lying outside the ϵ-decision boundary, leading to an overfitted model with poor performance on the test data. Among XGB and ANN, ANN shows better current density prediction on the test data, with the MAPE below 50% for both supported and unsupported cases. On the other hand, all the models show ideal performance (MAPE values close to zero) when predicting efficiency.

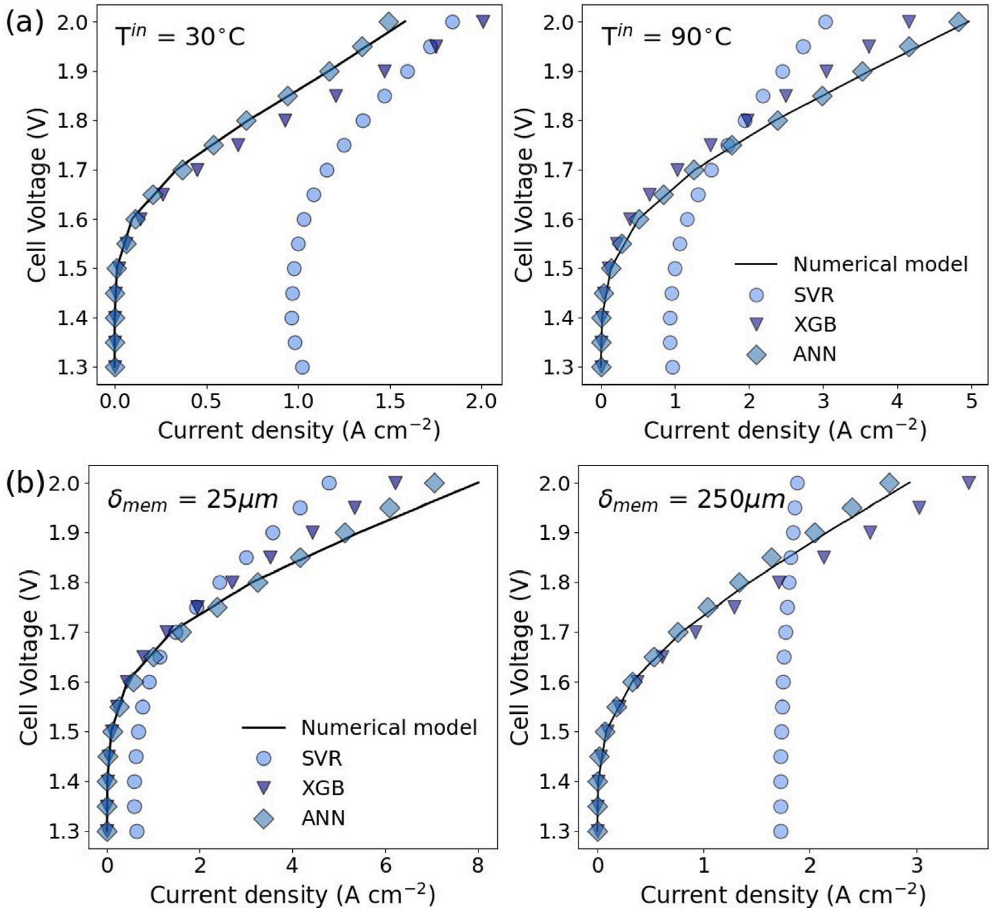

Fig. 6. Polarization curves generated by the developed data-driven models for two different values of (a) T^{in} and (b) δ_{mem}. These selected values that are located outside the range on which the models are trained from dataset 1.

Table 3
The mean absolute percentage error (MAPE) computed from the testing data set.

Supported				Unsupported		
Predictor	SVR	XGB	ANN	SVR	XGB	ANN
Current density	98.37×10^2	1.73×10^2	24.28	53.04×10^2	1.027×10^2	30.02
Efficiency	0.412	0.0027	0.0996	0.412	0.0019	0.0445

3.2.2. Model comparison

To further discern the differences between the developed models, we plot the polarization and efficiency curves at two random conditions and compare them with the corresponding curves generated from the 1D numerical model. It is to be noted that these data points were not a part of the original training and test datasets. The polarization curves for the supported catalyst case shown in Fig. 5(a) correspond to T^{in} = 80°, P_a^{in} = 1 bar, δ_{mem} = 127 μm, σ_a^{cl} = 350 S m^{-1}, δ_{cl}^a = 12 μm and δ_{ptl}^a = 400 μm. While all the three models show good agreement with the numerical model, the current densities predicted by the SVR model match well with the numerical model at voltages (1.7–1.9 V). For voltages below 1.7 V, we notice that the SVR model slightly overpredicts current density values. Conversely, it shows under prediction when the cell is operated at voltages over 1.9 V. However, all three models show good predictions in current density for supported catalysts as shown in Fig. 5(b). The corresponding operating conditions and transport properties are shown in Fig. 5. In line with our observations in Figs. 3 and 4 (bottom row), all three models show good comparison with the efficiency prediction of the CFD model for the supported and unsupported scenarios.

The two polarization curves exhibit high non-linearities as cell potential increases. As such, to make accurate predictions, a data-driven model must not only be able to capture this non-linear behavior, but also be robust enough to handle over-fitting. The XGB method uses an ensemble of decision trees which are trained on sequential residuals of the data. By considering the predictions of each of these decision trees, the XGB model handles non-linearities. This ensemble technique using negative gradient of the squared loss function also reduces the impact of outliers, thereby, circumventing over-fitting. Similarly, ANNs handle complex non-linear relationships between input features and output variables by learning patterns through layers of interconnected nodes. Each individual node consists of a non-linear rectified linear unit (ReLU) activation function. As the data transmits to-and-fro through each of these layers, the network continues to learn the intricate relationship between current density and the input variables listed in Table 2. While there are several approaches to intercept over-fitting [34], we have used the early stopping technique to prevent this. As such, the model gives accurate predictions well on unseen input conditions. On the other hand, the SVR generates an optimal hyperplane in the high-dimensional feature space. This hyperplane may not effectively capture the complex non-linear relationship. Additionally, the SVR method has fewer hyperparameters when compared to the ANN method. This reduces its flexibility to fine-tune the model and provide better predictions.

In general, data-driven models work well within the parameter space on which they have been trained. However, it is interesting to

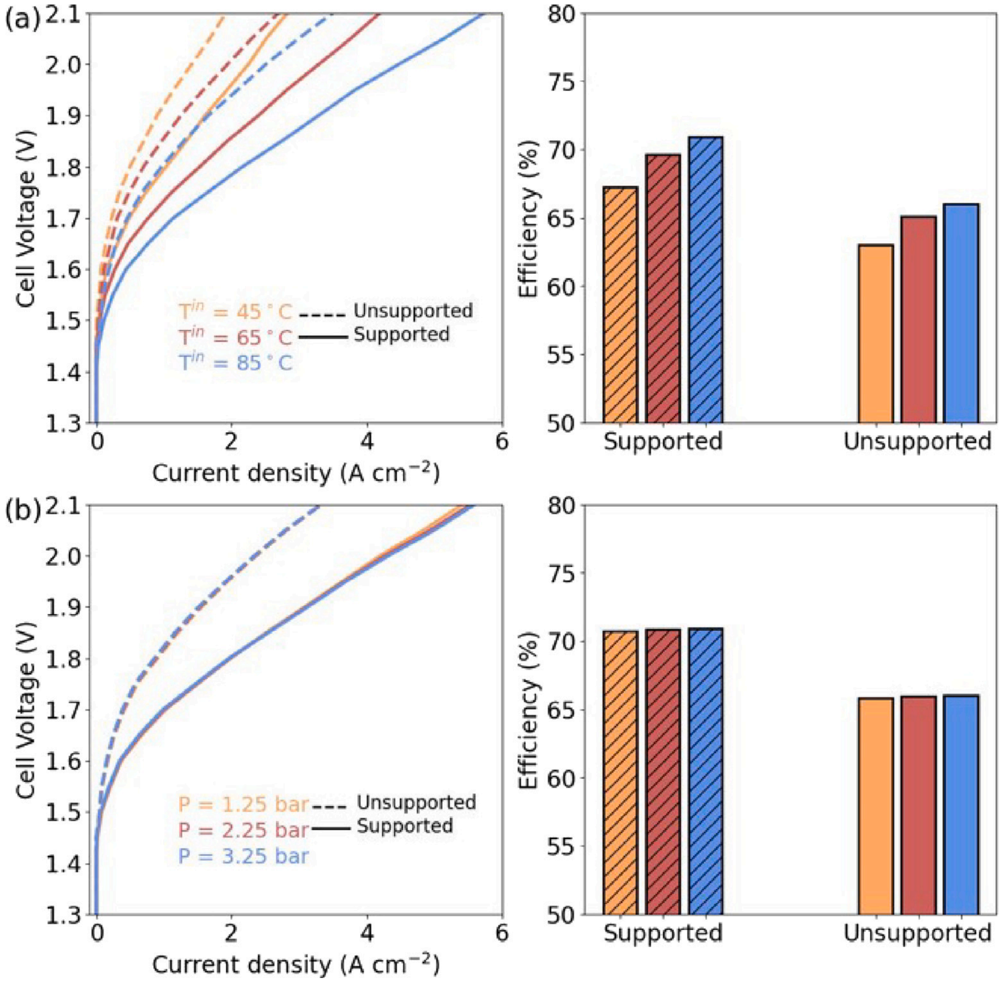

Fig. 7. Influence of operating conditions (a) Operating temperature (T^{in}) and (b) Anodic operating pressure (P_a^{in}) on polarization curves and cell efficiency. The cell efficiency is computed at a constant current density of I = 1.8 A cm^{-2}. The supported and unsupported catalyst layers are marked by solid and dashed lines, respectively.

probe into performance capabilities of these models when used for parameter values lying outside this training. To illustrate this effect, Fig. 6 shows the polarization curves obtained for the values of (a) T^{in} and (b) δ_{mem} extending outside the training range, as shown in Table 2. The other parameters are fixed with the underlined values. Fig. 6(a) illustrates the comparison between the three models when the operating temperature is extrapolated by 10 °C from the boundary values. We observe that the SVR model clearly fails to match the current densities predicted by the numerical model. For both $T^{i}n$ = 30 °C and 90 °C, the model shows a difference in current density of around 1 A/cm^2 when compared to that the current density given by the numerical model at the operating voltage of 1.3 V. The SVR model continues to over predict current density for the entire range of operating voltage at T^{in} = 30 °C. Whereas for T^{in} = 90 °C, the predictions of the model change from over to under prediction around the operating voltage of 1.7 V. Both XGB and ANN models show good comparison with the numerical model. While we notice a slight deviation in current density prediction by the XGB model at higher operating voltages, the prediction of the ANN model match well with the numerical model for the entire range of operating voltage. Apart from T^{in}, we have also varied the membrane thickness (δ_{mem}), to probe into the developed models capability to predict outside the range of their training data. While the trained input values for δ_{mem} varied from 50 μm to 180 μm, we selected 25 μm and 250 μm as test values to check model's extrapolation performance. This corresponds to a deviation in membrane thickness of 25 μm (very thin) and 70 μm (very thick) from the lower and upper boundary values, respectively. For the test cases, we notice from Fig. 6(b) that the ANN once again performs best, followed by the XGB model which shows a slight deviation at higher operating voltages. While the SVR model shows least performance, it is interesting to notice that while the model was capable to capturing the trend of the polarization curve given by the numerical for δ_{mem} = 25 μm, it completely fails when δ_{mem} = 250 μm. However, the ANN model with its multiple hidden layers and non-linear activation nodes, was capable of accurately learning the patterns and relationship between the input parameters and output variables to make correct extrapolations.

From the preceding sections, it is evident that both XGB and ANN models are capable of predicting IV-curves and efficiencies with high accuracy. Since XGB uses decision trees, it is more interpretable in comparison to ANN. Therefore, we select XGB for the discussion on feature importance in a later section and deploy the developed and calibrated ANN model to conduct parametric analysis in the next section.

3.2.3. Parametric analysis (ANN model)

The influence of different operating conditions and transport properties is discussed in this section for both supported and unsupported anode catalyst layers. The effect of these parameters is elucidated through polarization curves and cell efficiency (η_e) at a given current density of 1.8 A/cm^2.

(a) Operating conditions

Fig. 7 illustrates the effect of (a) operating temperature and (b) anodic operating pressure on the polarization curves and cell efficiency.

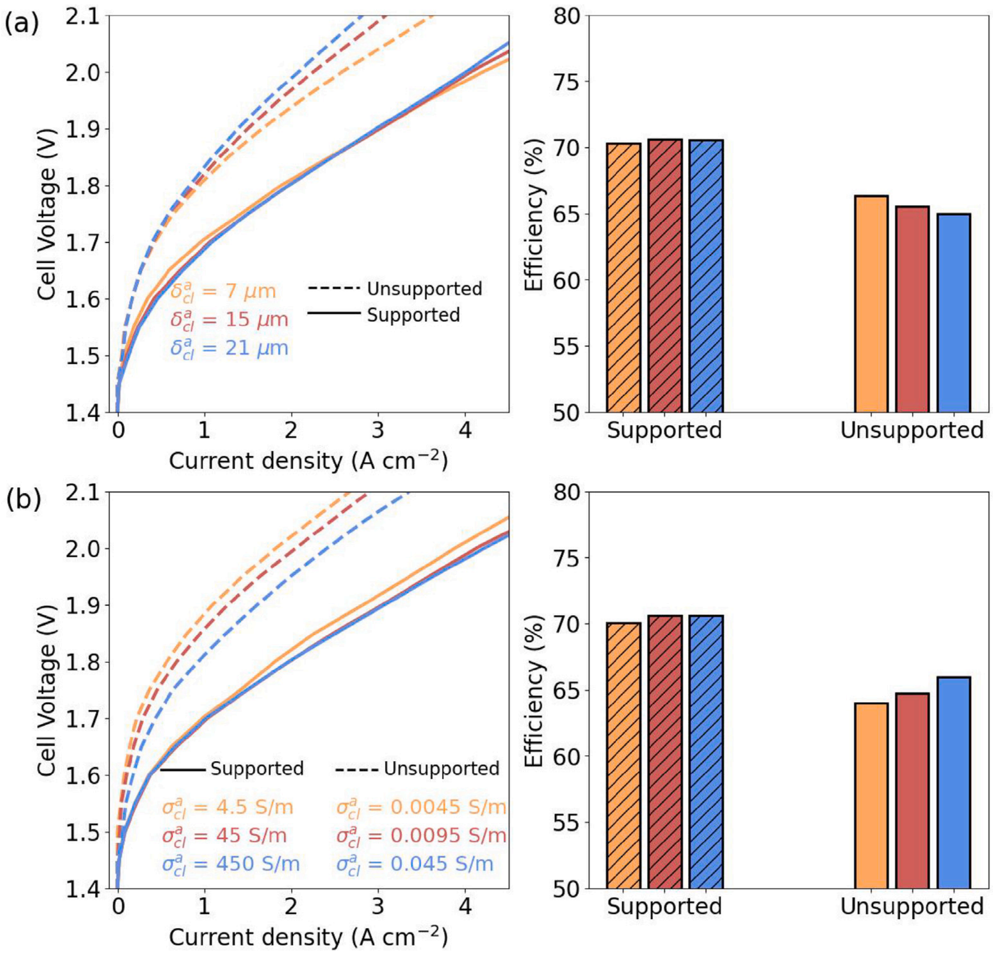

Fig. 8. Influence of anode catalyst transport properties such as (a) Thickness ((δ^a_{cl})) and (b) Electronic conductivity (σ^a_{cl}) on polarization curves and cell efficiency. The cell efficiency is computed at a constant current density of I = 1.8 A cm^{-2}. The supported and unsupported catalyst layers are marked by solid and dashed lines, respectively.

The solid lines and dashed lines represent supported and unsupported anode catalyst layers, respectively. As each of these parameters is varied, we have kept the values of other features constant, as shown by the underlined base parameter values in Table 2. For conducting parametric analysis, we have considered values which are slightly different from those used to train and test the model. However, it should be emphasized that the performance of a data-based model depends not only on the amount of data but also on the variety of data it has seen. Therefore, these models are capable of making accurate predictions when the values of the input features are within range of the feature values they were trained upon.

Fig. 7(a) illustrates the influence of T^{in} which is set to be 45°, 65°, and 85 °C. We observe that the current density increases with an increase in the operating temperature. This temperature-dependent enhancement in current density becomes more profound when the cell is operated at high voltages (crosses the kinetic regime). This behavior is attributed to the fact that as temperature increases, the electrochemical reaction rate and the protonic conductivity of the membrane increases, consequently, the proton transfer rate across the membrane increases [35]. This results in high current densities at elevated temperatures. In general, for a given temperature, we notice that supported catalyst layers show a higher value of current density when compared to unsupported catalyst layers. We also observe that the rate of increase in current density as temperature increases is larger for supported catalysts. Similarly, a 5.4% increase in η_e is noticed when T^{in} is increased from 45 °C to 85 °C for supported catalyst layers when compared to a 4.7% increment for the unsupported case.

A negligible influence of the anodic operating pressure on current density and cell efficiency is observed for both supported and unsupported catalyst layers. Fig. 7(b) shows that the polarization curves almost overlap each other as P^{in}_a is increased from 1.25 bar to 3.25 bar. Similar behavior was exhibited from the experimental results [36,37] when anode pressure is varied between 1 bar to 6 bar.

(b) Transport properties

We next discuss the effect of transport properties at the anode like the catalyst layer thickness, electronic conductivity, and porous transport layer thickness on cell performance. Investigation of anode side transport properties is crucial as the water splitting, which is the rate limiting step, takes place on this side of the cell. The kinetics of OER are more sluggish compared to the HER.

The effect of (δ^a_{cl}) for supported and unsupported catalyst layers is shown in Fig. 8(a). For the range of δ^a_{cl} considered in this work, we notice that it has no significant influence on the supported catalyst. The cell efficiency follows a similar trend for this type of catalyst layers. A dependence on δ^a_{cl} can be spotted in Fig. 8(a) for unsupported catalysts, especially at cell voltages above 1.8 V. We observe that at these higher voltages, the current density increases as δ^a_{cl} decreases. This inverse relationship between current density and δ^a_{cl} can be explained by the fact that a thinner δ^a_{cl} reduces mass transport limitations and ohmic losses, especially at higher current densities. The cell efficiency for unsupported catalyst layers computed at 1.8 A/cm^2 is in line with this observation.

From the preceding discussion, we observed that the influence of δ^a_{cl} is different on supported and unsupported catalyst layers. We now

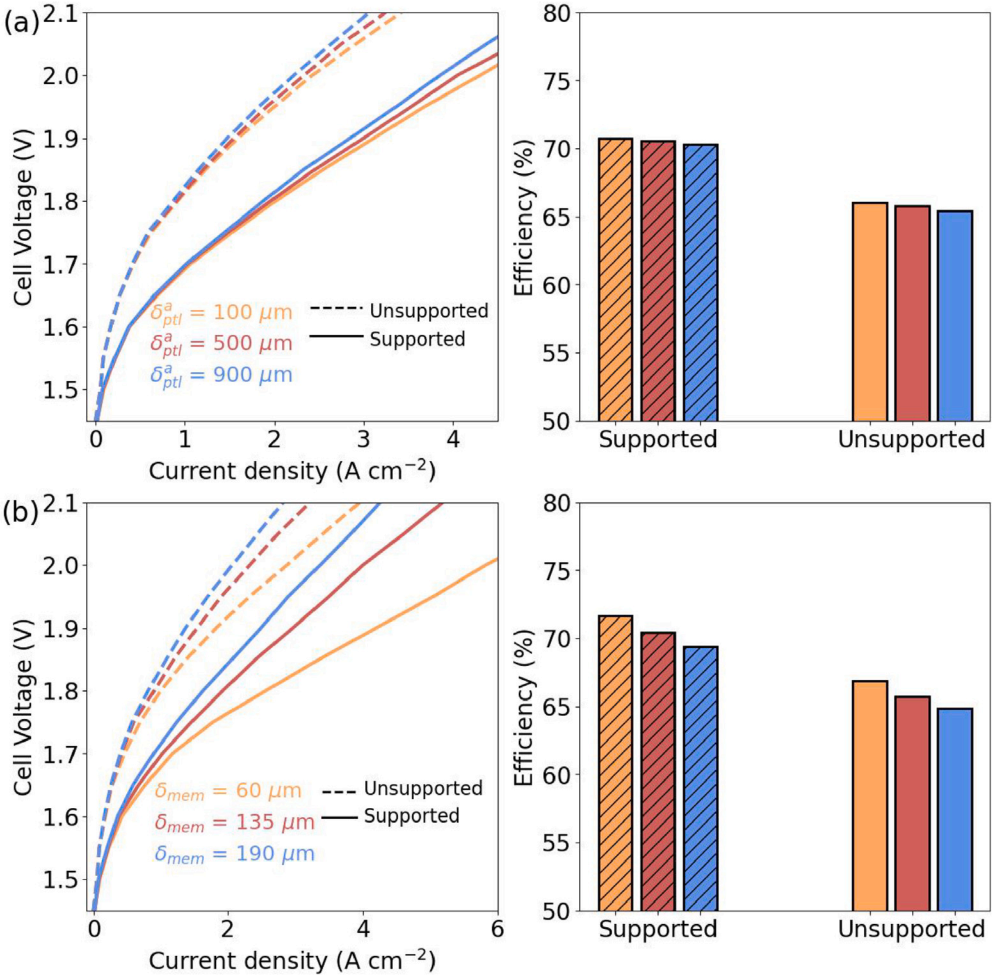

Fig. 9. Influence of transport properties such as (a) anode porous transport layer thickness (δ^a_{ptl}) and (b) Membrane thickness ((δ_{mem})) on polarization curves and cell efficiency. The cell efficiency is computed at a constant current density of I = 1.8 A cm^{-2}. The supported and unsupported catalyst layers are marked by solid and dashed lines, respectively.

investigate the role of the electronic conductivity of the catalyst layer which results in such differential behavior. The σ^a_{cl} is an important parameter which governs the current density. As σ^a_{cl} is varied from 4.5×10^{-3} S m^{-1} to 4.5×10^{-2} S m^{-1} for unsupported catalysts, the electron transfer rate increases as they face lesser ohmic resistance with increasing electronic conductivity. This leads to an increase in the current density with an increase in σ^a_{cl} as shown in Fig. 8(b). Consequentially, a 3.1% increase in cell efficiency in noted at I = 1.8 A/cm^{-2}. In contrast, a 0.7% increase in cell efficiency is observed for supported catalysts. An increase in σ^a_{cl} has no significant effect on the current density-voltage relationship, which signifies that σ^a_{cl} acts as a non-limiting factor in determining current density for supported catalysts. In general, when the catalyst layers are unsupported, the effective electronic conductivity is significantly reduced owing to the interference of the ionomer network [23].

For both supported and unsupported catalyst layers, with a fixed electronic conductivity of 1250 S/m, we do not observe a significant dependence of current density on δ^a_{ptl} when it is varied from 100 μm to 900 μm, particularly at lower cell voltages (<1.8 V). However, at higher cell voltages we discern that a thinner δ^a_{ptl} promotes higher current density.

The effect of membrane thickness on current density and efficiency is presented in Fig. 9(b) for δ_{mem} = 60 μm, 135 μm and 190 μm. We observe that the current density decreases as δ_{mem} increases. This inverse relationship between membrane thickness and current density is related to the lower transport resistance offered by a thinner membrane. Thicker membranes increase the ohmic resistance by providing a longer path to the flow of H$^+$ ions across the membrane [38]. Therefore, fewer hydrogen ions pass through the membrane and it results in the decrease of the overall electrochemical reaction rate. However, a thinner membrane could also lead to undesirable gas crossover [39] and parasitic reactions.

While data-based models are known for their black-box nature, XGB offers a higher degree of explainability when compared to ANNs [40]. Each decision tree of the XGB model is interpretable as it is split based on the importance of a particular feature. The splitting process is accomplished by minimizing the impurity in the child nodes, thereby, determining the importance of each feature. On the other hand, ANNs are complex to interpret as their predictions are based on the distributed nature in which the information is transferred through inter-connected nodes. Therefore, we use the XGB model to investigate the importance of input features in predicting current distribution in the following section.

3.2.4. Feature importance and sensitivity analysis (XGB model)

Fig. 10 lists the important features affecting the current density predictions by the XGB model for (a) supported and (b) unsupported catalyst layers. SHAP (SHapely Additive exPlanations) [41] is an approach based on cooperative game theory which provides information on how much "gain" (SHAP value) a particular feature contributes to the model's prediction. The SHAP summary plot illustrated in Fig. 10 not only ranks the important features but also gives us insight into how each of these parameters affected the current density predicted by the XGB model. For both supported and unsupported catalyst layers,

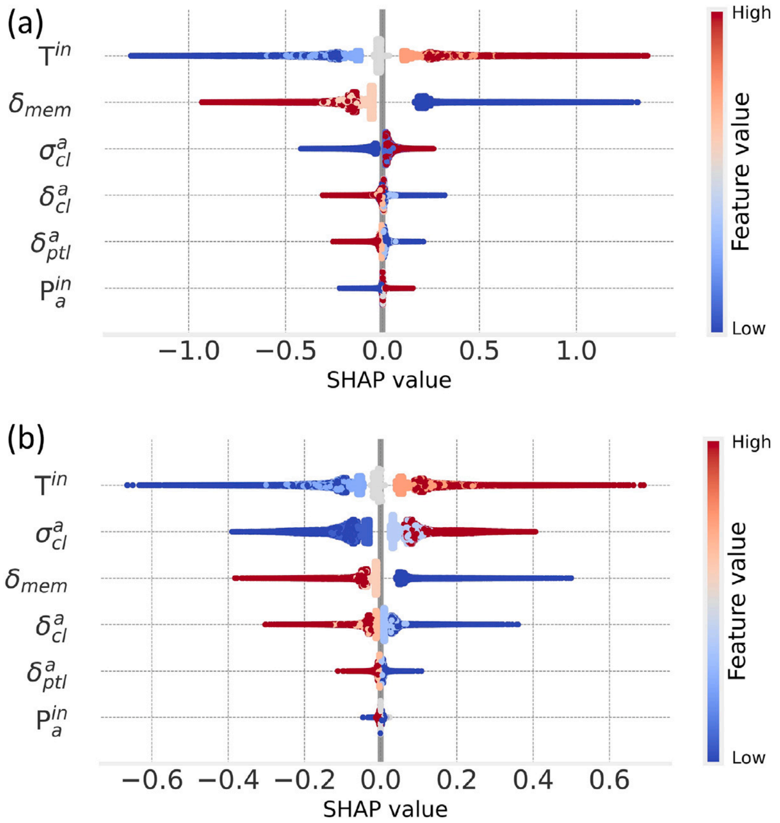

Fig. 10. Shapely additive explanations (SHAP) summary plots illustrating the relevant features and their influence in predicting current density for (a) Supported and (b) Unsupported catalyst layers. The XGB model is deployed to compute the SHAP values.

operating temperature is the most influencing parameter. We clearly observe that larger T^{in} values lead to higher current density as they have positive SHAP values. The other prominent factors influencing the current density are the membrane thickness, anode catalyst electronic conductivity, and anode catalyst layer thickness. A similar observation was observed in a recent work [5] on sensitivity analysis of PEMEC. It was reported that anode kinetics and membrane thickness are critical parameters governing the performance of PEMEC. From Fig. 10, we deduce that δ_{mem} and δ^a_{cl} form a negative relationship with the current density. This implies that increasing these parameters will result in the reduction of current density. Input parameters such as T^{in} and σ^a_{cl} shows a positive relationship as illustrated from the corresponding SHAP values for both types of catalyst layers. An interesting observation is noted in the ranking of δ_{mem} and σ^a_{cl} for supported and unsupported catalyst layers. We observe that for supported catalyst layers, δ_{mem} has a higher influence on the model's prediction over σ^a_{cl}. However, in the case of unsupported catalysts, the effect of σ^a_{cl} is greater than δ_{mem}. From the preceding section, we have seen that for supported catalyst layers σ^a_{cl} acts as a non-limiting factor, and increase in its value does not have a significant impact on the I-V relationship. Additionally, in case of supported catalyst layers, the support material becomes the primary pathway for electron transport. Consequently, the role of catalyst particles as a medium of electronic transport is reduced. These catalyst particles deposited on a support material have a higher electrochemically active surface area (ECSA) and better catalyst utilization resulting in higher catalytic activity per weight unit/mg of catalyst material. As such, they can be a promising pathway to reducing the overall system cost by decreasing the amount of deployed catalyst material. Since, the support material is the primary pathway for electronic transport, the SHAP results showcase δ_{mem} as a parameter of higher importance when compared to σ^a_{cl} for supported catalysts. In the case of unsupported catalyst layers, the contact between the catalyst particles is the primary pathway of electronic contact. Therefore, the model's predictions are more sensitive to the σ^a_{cl} than δ_{mem} in this scenario.

Table 4
Uncertainty range and probability distribution functions (PDF) of the input parameters, taken into account for the global SA.

Parameter	Unit	Uncertainty range	Distribution
T^{op}	°C	[40, 80]	$\mathcal{U}(40, 80)$
P^a	bar	[1, 3]	$\mathcal{U}(1, 3)$
δ_{cl}	μm	[8, 20]	$\mathcal{U}(8, 20)$
δ_{ptl}	μm	[200, 800]	$\mathcal{U}(200, 800)$
δ_{mem}	μm	[50, 180]	$\mathcal{U}(50, 180)$
σ^a_{cl} (supported catalyst)	S m^{-1}	[3.5, 350]	$\mathcal{U}(3.5, 350)$
σ^a_{cl} (unsupported catalyst)	S m^{-1}	[0.0035, 0.085]	$\mathcal{U}(0.0035, 0.085)$

We have also performed a sensitivity analysis (SA) study (Fig. 11) to gain a deeper understanding of the interdependence between various input parameters related to operating conditions and transport properties, and their resulting impact on the polarization behavior of PEMECs. We treat these input parameters as independent random variables, assigning them a uniform probability density function (PDF), as detailed in Table 4. Fig. 11 shows the first-order and total-order sensitivity indices (SIs) for the averaged current density for both supported and unsupported catalysts.

At this point, it is essential to clarify the difference between first-order and total-order SIs. The total-order SIs account for both the sensitivity due to first-order effects and the sensitivity resulting from interactions between a specific parameter and all other parameters. The sum of the total-order SIs is equal to or greater than one. If no higher-order interactions are present, the sum of the first-order and total-order SIs will equal one. From Fig. 11, we observe differences between the first-order and total-order SIs for some input parameters,

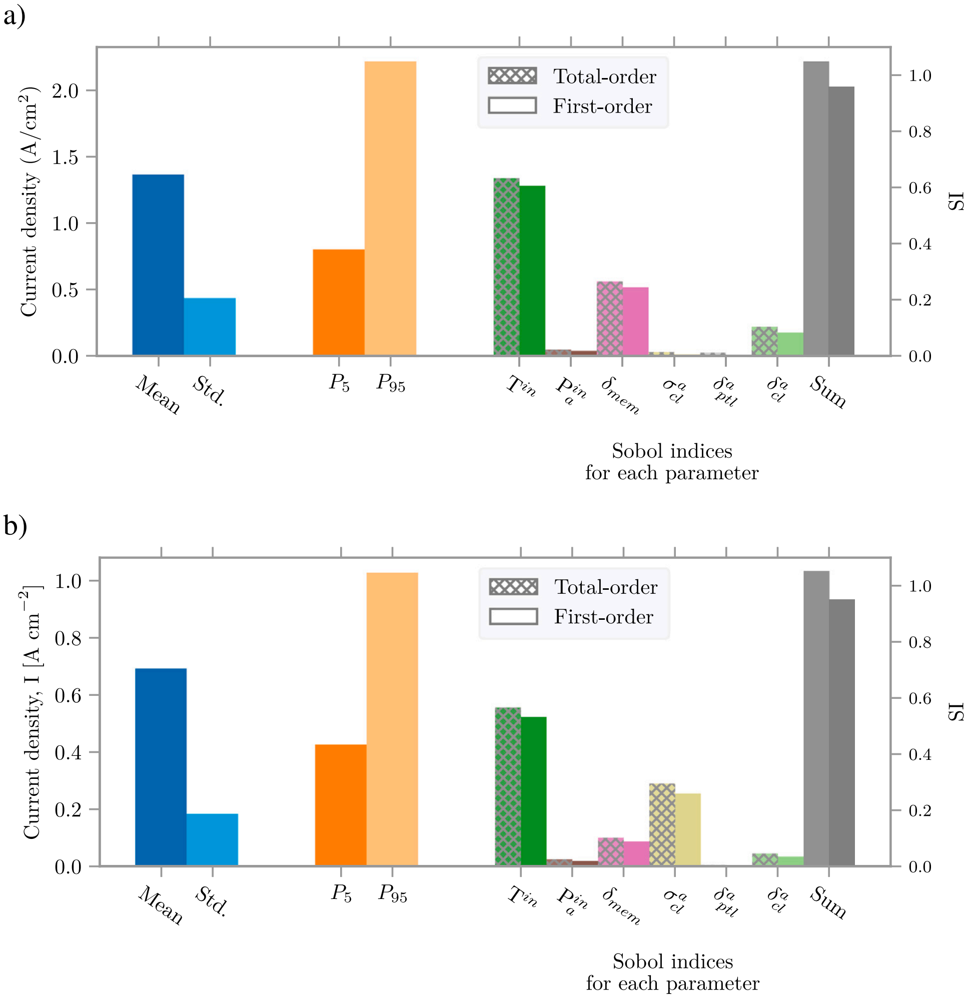

Fig. 11. The sensitivity for different features of the 1D physics-based PEMEC model charaterised by the mean, standard deviation and first-order Sobol indices of the averaged potential for (a) supported and (b) unsupported catalysts. The XGB model is used to perform the sensitivity analysis.

particularly in their sums. These differences might indicate that higher-order interactions have a significant influence on the polarization behavior.

From Fig. 11(a) and in the case of a supported catalyst, the most influential parameter for the polarization behavior after the operating temperature is the PEM layer thickness. This finding was also reported in our earlier study [5]. Another important finding of this analysis is the high impact of the anode catalyst layer thickness. This is in contrast with the SHAP analysis, but we need to emphasize here that the anode catalyst layer thickness might be strongly correlated with the effective electronic conductivity of the anode catalyst layer.This can be attributed to the fact that SA assumes each input feature as an independent parameter and does account to correlated parameters. From Fig. 11(b) and for the case of an unsupported catalyst, the most influential parameters are the operating temperature, the anode catalyst electronic conductivity, and the membrane and anode catalyst thickness. This is also confirmed by the SHAP plot in Fig. 10. From global SA, we observe the anode operating pressure and PTL thickness do not influence much the current density. The SHAP analysis has also ranked them in the last two positions.

4. Conclusions

The influence of transport properties on the performance of PEMEC with supported and unsupported catalyst layers is predicted through three different machine-learning methods: SVR, XGB, and ANN. The data required to train these models is generated from an analytical model and a one-dimensional, non-isothermal, two-phase flow numerical model of PEMEC. The input features comprised parameters related to operating conditions and transport properties of the MEA, in particular for the anode side. Statistical error metrics such as the MAE, RMSE, and R^2 were used to compare the predictive performance between these three models on the training and test datasets. Upon comparison, the ANN method is selected to conduct parametric analysis, and XGB model was considered for performing sensitivity analysis for PEMEC performance. The main conclusions are drawn as follows:

(1) When moving from an analytical model to a complex one-dimensional physics-based model, there is a significant increase in the number of parameters and computational cost. However, the black-box nature of ML models does not lead to such a significant increase in the training cost in order to provide accurate predictions.
(2) Among the three tested models, the SVR model has shown the least predictive performance on both the test datasets. This implies that the hyperplane generated by the SVR model could not capture the non-linearities, especially at higher voltages. Both XGB and ANN performed well in predicting current density and cell efficiency.
(3) In general, operating temperature has been shown to influence the current density positively, while a decrease in membrane

thickness leads to increasing current density. Furthermore, we observe that for unsupported catalyst layers, the anode catalyst conductivity and thickness influence the performance of PEMEC. However, for supported catalysts, variation in these parameters does not show any significant impact on PEMEC performance.

(4) Feature importance study and sensitivity analysis revealed that the operating temperature is the most important parameter affecting the cell performance. This is followed by membrane thickness and catalyst conductivity for supported catalyst, and vice-versa for unsupported catalyst layers.

The current study sheds light into deploying machine learning techniques, developed from data obtained from physics-based numerical simulations, to model PEMEC. This is in contrast to the predominant work done in using ML at an atomistic scale for catalyst screening from DFT-generated databases. Further exploration in optimizing key components of PEMEC and cell stack using principles of ML needs to be undertaken in future works. Furthermore, the approach of combining data driven models with sensitivity analysis can be a promising approach in real-time operation of PEM electrolyzer at an industrial scale.

At this point, it is important to mention some limitations in using only synthetic data obtained from physics-based numerical simulations. Numerical simulations often rely on idealized boundary conditions, material properties and simplifications of the governing equations. Therefore, practical issues such as material impurities or unexpected side reactions may not be totally captured by physics-based synthetic data. This may result to inadequate model generalization in real-world conditions. Along this direction, in the future, we aim to improve the generalization of the models generated by the approach of this work by adding experimental data from cell experiments. This will enhance the robustness and accuracy of the developed surrogate models presented in our work.

CRediT authorship contribution statement

K. Ashoke Raman: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Linus Hammacher:** Writing – original draft, Data curation. **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. **Violeta Karyofylli:** Writing – review & editing, Validation, Investigation, Formal analysis.

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.

Appendix A. The effect of training and test data split ratio

In general, the model should have sufficient data to learn the different patterns and varying relationships between the input parameters and the target variables of the model. To illustrate the effect of split ratio (SR), we have selected SVR and XGB models for dataset 2 and investigate their effect on predicting the current density on the test data. Table 5 presents the corresponding performance metrics. Two

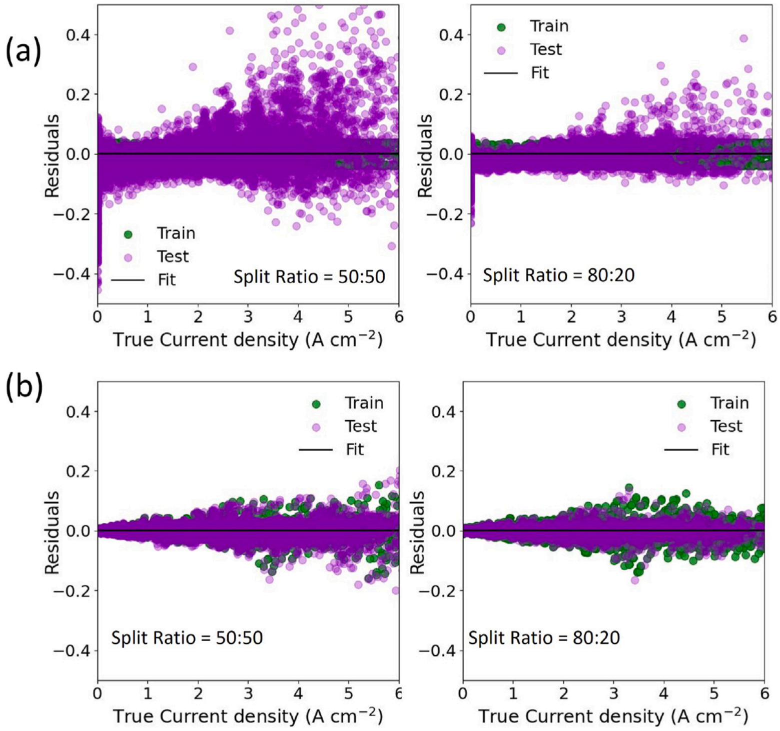

Fig. 12. Residual plots showing the comparison in predicting current density for (a) SVR and (b) XGB models for two different data split ratios (SR). The plots shown correspond to the model developed from the 1D physics based model.

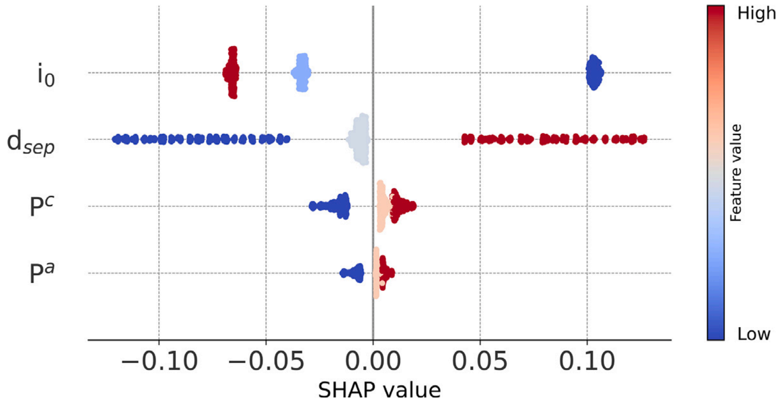

Fig. 13. Shapely additive explanations (SHAP) summary plots illustrating the relevant features and their influence in predicting cell voltage. The plots shown correspond to the model developed from the 0D semi empirical model.

Table 5
The error metrics computed from the test part of the dataset 2 for different split ratios.

SVR				XGB		
Split ratio	RMSE	R^2	MAPE	RMSE	R^2	MAPE
80:20	0.046	0.999	98.37×10^2 %	0.0140	0.999	1.73×10^2 %
70:30	0.054	0.998	111.79×10^2 %	0.0137	0.999	1.75×10^2 %
60:40	0.0646	0.998	125.86×10^2 %	0.0147	0.999	1.68×10^2 %
50:50	0.08	0.997	146.07×10^2 %	0.0164	0.999	1.66×10^2 %

distinct observations can be made from the error metrics shown in Table 5. Firstly, we notice that the both RMSE and MAPE are increased for the SVR model when SR is varied from 80:20 to 50:50. Secondly, this effect of changing SR does not have a significant influence on the model performance. This implies that the choice of SR is related to how well the ML model learns from a given set of training data. Fig. 12 further elucidates this observation through the residual plots for the (a) SVR and (b) XGB model.

Appendix B. Feature importance for dataset 1

Fig. 13 ranks the important parameters effecting the cell voltage based on their SHAP values. We observe that the exchange current density is the most important parameter influencing the cell voltage. This is followed by the membrane thickness. When compared to these two parameter, both the anodic and cathodic pressure do not show significant influence in predicting cell voltage.

Data availability

Data will be made available on request.

References

[1] Noyan Omer Faruk, Hasan Muhammad Mahmudul, Pala Nezih. A global review of the hydrogen energy eco-system. Energies 2023;16(3):1484.
[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] Shirvanian Paige, van Berkel Frans. Novel components in proton exchange membrane (PEM) water electrolyzers (PEMWE): Status, challenges and future needs. A mini review. Electrochem Commun 2020;114:106704.
[4] Khatib Fawwad Nisar, Wilberforce Tabbi, Ijaodola Oluwatosin, Ogungbemi Emmanuel, El-Hassan Zaki, Durrant Andy, Thompson James, Olabi Abdul Ghani. Material degradation of components in polymer electrolyte membrane (PEM) electrolytic cell and mitigation mechanisms: A review. Renew Sustain Energy Rev 2019;111:1–14.
[5] Karyofylli Violeta, Danner Yannik, Raman K Ashoke, Kungl Hans, Karl André, Jodat Eva, Eichel Rüdiger-A. Sensitivity analysis and uncertainty quantification in predictive modeling of proton-exchange membrane electrolytic cells. J Power Sources 2024;600:234209.
[6] Sutharssan Thamo, Montalvao Diogo, Chen Yong Kang, Wang Wen-Chung, Pisac Claudia, Elemara Hakim. A review on prognostics and health monitoring of proton exchange membrane fuel cell. Renew Sustain Energy Rev 2017;75:440–50.
[7] 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.
[8] 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.
[9] 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.
[10] Li Hao, Xu Shaopeng, Wang Min, Chen Ziheng, Ji Fengfeng, Cheng Kewei, Gao Zhengyang, Ding Zhao, Yang Weijie. Computational design of (100) alloy surfaces for the hydrogen evolution reaction. J Mater Chem A 2020;8(35):17987–97.
[11] Ran Nian, Sun Bo, Qiu Wujie, Song Erhong, Chen Tingwei, Liu Jianjun. Identifying metallic transition-metal dichalcogenides for hydrogen evolution through multilevel high-throughput calculations and machine learning. J Phys Chem Lett 2021;12(8):2102–11.
[12] García-Salaberri Pablo A. 1D two-phase, non-isothermal modeling of a proton exchange membrane water electrolyzer: An optimization perspective. J Power Sources 2022;521:230915.
[13] 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).
[14] 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.
[15] Ding Rui, Chen Yawen, Rui Zhiyan, Hua Kang, Wu Yongkang, Li Xiaoke, Duan Xiao, Wang Xuebin, Li Jia, Liu Jianguo. Guiding the optimization of membrane electrode assembly in a proton exchange membrane water electrolyzer by machine learning modeling and black-box interpretation. ACS Sustain Chem Eng 2022;10(14):4561–78.
[16] Mohamed Amira, Ibrahem Hatem, Yang Rui, Kim Kibum. Optimization of proton exchange membrane electrolyzer cell design using machine learning. Energies 2022;15(18):6657.
[17] Nafeh Abd El-Shafy A. Hydrogen production from a PV/PEM electrolyzer system using a neural-network-based MPPT algorithm. Int J Numer Modelling, Electron Netw Devices Fields 2011;24(3):282–97.
[18] 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.
[19] 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.
[20] Schalenbach Maximilian, Tjarks Geert, Carmo Marcelo, Lueke Wiebke, Mueller Martin, Stolten Detlef. Acidic or alkaline? Towards a new perspective on the efficiency of water electrolysis. J Electrochem Soc 2016;163(11):F3197.

[21] 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.

[22] Siracusano S, Baglio V, D'Urso C, Antonucci V, Aricò AS. Preparation and characterization of titanium suboxides as conductive supports of IrO2 electrocatalysts for application in SPE electrolysers. Electrochim Acta 2009;54(26):6292–9.

[23] Mandal Manas, Moore Michael, Secanell Marc. Measurement of the protonic and electronic conductivities of PEM water electrolyzer electrodes. ACS Appl Mater & Interfaces 2020;12(44):49549–62.

[24] Yekkehkhany Bahareh, Safari Abdolreza, Homayouni Saeid, Hasanlou Mahdi. A comparison study of different kernel functions for SVM-based classification of multi-temporal polarimetry SAR data. Int Arch Photogramm Remote Sens Spat Inf Sci 2014;40:281–5.

[25] Chanklan Ratiporn, Kaoungku Nuntawut, Suksut Keerachart, Kerdprasop Kittisak, Kerdprasop Nittaya. Runoff prediction with a combined artificial neural network and support vector regression. Int J Mach Learn Comput 2018;8(1):39–43.

[26] Goodfellow Ian, Bengio Yoshua, Courville Aaron. Deep learning. MIT Press; 2016, http://www.deeplearningbook.org.

[27] Mänken Christian Frederik, Uecker Jan, Schäfer Dominik, De Haart LGJ, Eichel Rüdiger-A. Impact of electrochemical impedance spectroscopy dataset curation on solid oxide cell degradation assessment. J Electrochem Soc 2024.

[28] Li Hong-Wei, Qiao Bin-Xin, Liu Jun-Nan, Yang Yue, Fan Wenxuan, Lu Guo-Long. A data-driven framework for performance prediction and parameter optimization of a proton exchange membrane fuel cell. Energy Convers Manage 2022;271:116338.

[29] Li Hong-Wei, Xu Bo-Shi, Du Chang-He, Yang Yue. Performance prediction and power density maximization of a proton exchange membrane fuel cell based on deep belief network. J Power Sources 2020;461:228154.

[30] Chandesris M, Médeau V, Guillet N, Chelghoum S, Thoby D, Fouda-Onana F. Membrane degradation in PEM water electrolyzer: Numerical modeling and experimental evidence of the influence of temperature and current density. Int J Hydrog Energy 2015;40(3):1353–66.

[31] Debe MK, Hendricks SM, Vernstrom GD, Meyers M, Brostrom M, Stephens M, Chan Q, Willey J, Hamden M, Mittelsteadt CK, et al. Initial performance and durability of ultra-low loaded NSTF electrodes for PEM electrolyzers. J Electrochem Soc 2012;159(6):K165.

[32] Lamy Claude, Millet Pierre. A critical review on the definitions used to calculate the energy efficiency coefficients of water electrolysis cells working under near ambient temperature conditions. J Power Sources 2020;447:227350.

[33] Chen Tianqi, Guestrin Carlos. 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.

[34] Legala Adithya, Zhao Jian, Li Xianguo. Machine learning modeling for proton exchange membrane fuel cell performance. Energy AI 2022;10:100183.

[35] Li Hua, Inada Akiko, Fujigaya Tsuyohiko, Nakajima Hironori, Sasaki Kazunari, Ito Kohei. Effects of operating conditions on performance of high-temperature polymer electrolyte water electrolyzer. J Power Sources 2016;318:192–9.

[36] Garcia-Navarro Julio Cesar, Schulze Mathias, Friedrich Kaspar Andreas. Measuring and modeling mass transport losses in proton exchange membrane water electrolyzers using electrochemical impedance spectroscopy. J Power Sources 2019;431:189–204.

[37] Ojong Emile Tabu, Kwan Jason Tai Hong, Nouri-Khorasani Amin, Bonakdarpour Arman, Wilkinson David P, Smolinka Tom. Development of an experimentally validated semi-empirical fully-coupled performance model of a PEM electrolysis cell with a 3-D structured porous transport layer. Int J Hydrog Energy 2017;42(41):25831–47.

[38] Toghyani Somayeh, Fakhradini Soheil, Afshari Ebrahim, Baniasadi Ehsan, Jamalabadi Mohammad Yaghoub Abdollahzadeh, Shadloo Mostafa Safdari. Optimization of operating parameters of a polymer exchange membrane electrolyzer. Int J Hydrog Energy 2019;44(13):6403–14.

[39] Trinke P, Bensmann B, Hanke-Rauschenbach R. Experimental evidence of increasing oxygen crossover with increasing current density during PEM water electrolysis. Electrochem Commun 2017;82:98–102.

[40] Louppe Gilles. Understanding random forests: From theory to practice. 2014, arXiv preprint arXiv:1407.7502.

[41] 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.