

Original Article

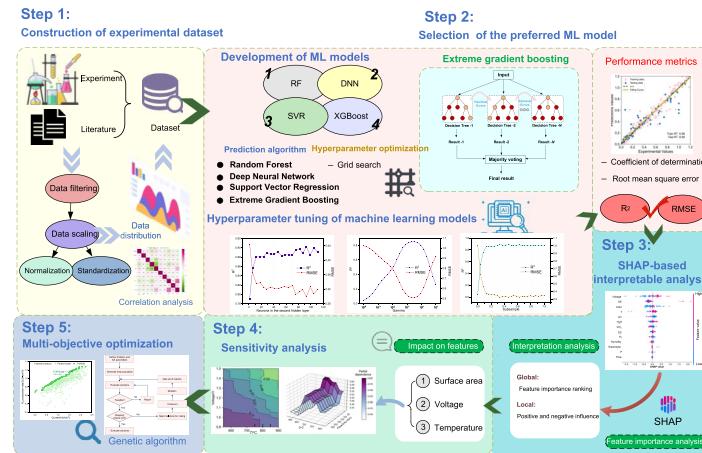
Machine learning-assisted prediction and optimization of solid oxide electrolysis cell for green hydrogen production

Qingchun Yang ^{a,b,*}, Lei Zhao ^a, Jingxuan Xiao ^c, Rongdong Wen ^a, Fu Zhang ^b, Dawei Zhang ^{a,**}^a School of Chemistry and Chemical Engineering, Hefei University of Technology, Hefei, 230009, China^b East China Engineering Science and Technology Co. Ltd, Hefei, 230011, China^c Faculty of Innovation Engineering, Macau University of Science and Technology, 999078, China

HIGHLIGHTS

- Four machine learning models are developed to predict the Ohmic resistance, current density, and H₂ production rate of the SOEC process.
- An interpretable analysis based on the SHAP method is performed to explain the predictive performance of the preferred XGBoost model.
- Interaction relationship between the input parameters and various output indicators is investigated by the PDP analysis.
- XGBoost model is coupled with the GA algorithm to achieve the maximum H₂ production rate with the lowest current in the SOEC process.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Solid oxide electrolysis cell
Machine learning
H₂ production rate
Feature importance analysis
System optimization

ABSTRACT

The solid oxide electrolysis cell (SOEC) holds great promise to efficiently convert renewable energy into hydrogen. However, traditional modeling methods are limited to a specific or reported SOEC system. Therefore, four machine learning models are developed to predict the performance of SOEC processes of various types, operating parameters, and feed conditions. The impact of these features on the SOECs outputs is explained by the Shapley additive explanations and partial dependency plot analyses. The preferred model is integrated with a genetic algorithm to determine the optimal values of each input feature. Results show the improved extreme gradient enhanced regression (XGBoost) algorithm is the core of the machine learning model of the process since it has the highest R² (> 0.95) in the three outputs. The electrolytic cell descriptors have a greater impact on the system performance, contributing up to 54.5%. The effective area, voltage, and temperature are the three most influential factors in the SOEC system, contributing 21.6%, 16.6%, and 13.0% to its performance. High temperature, high pressure, and low effective area are the most favorable conditions for H₂ production rate. After conducting multi-objective optimization, the optimal current intensity and hydrogen production rate were determined to be 1.61 A/cm² and 1.174 L/(h·cm²).

* Corresponding author. School of Chemistry and Chemical Engineering, Hefei University of Technology, Hefei, 230009, China.

** Corresponding author.

E-mail addresses: ceqcyang@hfut.edu.cn (Q. Yang), zhangdw@ustc.edu.cn (D. Zhang).<https://doi.org/10.1016/j.gce.2024.04.004>

Received 20 March 2024; Received in revised form 17 April 2024; Accepted 29 April 2024

2666-9528/© 2024 Institute of Process Engineering, Chinese Academy of Sciences. Publishing services by Elsevier B.V. on behalf of KeAi Communication Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Nomenclature

X	x variable sample values
\bar{X}	x variable mean values
Y	y variable sample values
\bar{Y}	y variable mean values
x_i^*	Sample standardized values
x_i^n	Sample true values
μ_i	Feature average value
σ_i	Feature standard deviation
n	Samples total
k	Samples index
y_k	Experimental values
\hat{y}_k	Predicted values
mean(y)	Corresponding experimental average value
f_{base}	Model baseline
f_{H_2}	Hydrogen production rate
f_i	Current density
w_1	Weight factor

Abbreviations

SOEC	Solid oxide electrolysis cell
------	-------------------------------

ALK	Alkaline water
PEM	Proton exchange membrane
RF	Random forest
DNN	Deep neural network
SVR	Support vector regression
XGBoost	Extreme gradient boosting
EC	Electrolytic cell
OP	Operation parameters
FC	Feed conditions
ML	Machine learning
GA	Genetic algorithm
ET	Electrolyte thickness
CT	Cathode thickness
RMSE	Root mean square error
PCC	Pearson correlation coefficient
R ²	Coefficient of determination
SHAP	Shapley additive explaining
PDP	Partial dependence plots
LSM	Lanthanum strontium manganite
YSZ	Yttria stabilized zirconia

1. Introduction

The extensive consumption of conventional fossil fuels has caused serious problems of energy crisis, ecological destruction, and climate change [1]. To address the climate issues caused by greenhouse gas emissions and alleviate dependence on traditional fossil fuels, various countries are actively taking measures to exploit and utilize renewable energy resources (such as solar, wind, and hydro) for promoting their energy transformation [2]. Green hydrogen is such a renewable energy utilization technology, which is produced through the electrolysis of water using renewable energy sources. It has gradually become a key support for the large-scale, diversified, and high-quality development of hydrogen energy under the dual drive of energy transformation and carbon reduction.

The alkaline water (ALK) electrolysis, proton exchange membrane (PEM) water electrolysis, and solid oxide electrolysis cell (SOEC) technologies are the three main methods for green hydrogen production [3]. The basic principles and characteristics of these three technologies are compared in Fig. 1. Compared to the previous two technologies, SOEC does not require expensive catalysts and has faster reaction kinetics, higher efficiency, and lower electrolysis voltage as shown in Fig. 1b [4].

Moreover, it can integrate with the waste heat from industrial processes, thus improving energy utilization efficiency and decreasing electricity consumption [5]. For example, Penchini et al. [6] found that SOEC technology can save more than 30% of electricity consumption compared to ALK and PEM when it is coupled with low-grade waste heat such as steel metallurgy, chemical industry, and nuclear power. In addition, the SOEC system adopts a modular design, which is easy to install and can flexibly adjust the hydrogen production scale according to actual needs. Hence, the hundred-kilowatt-scale SOEC has achieved preliminary commercialization in many countries (such as the United States, Germany, and Denmark) for its promising development prospects [7].

Although SOEC technology has shown great potential for development, its production cost is the highest in hydrogen production by electrolysis of water [8]. As reported by the International Energy Agency (IEA) [9], the replacement cost of the SOEC is second only to the cost of electricity input, accounting for nearly 40% of its total hydrogen production cost. Thus, it is urgent to optimize its hydrogen production efficiency and service life for wide promotion. Facing the high capital cost and fragile electrode of SOEC, it would be extremely challenging to investigate the operability issues using an actual SOEC stack [10]. Alternatively, the modeling and simulation approach is regarded as one

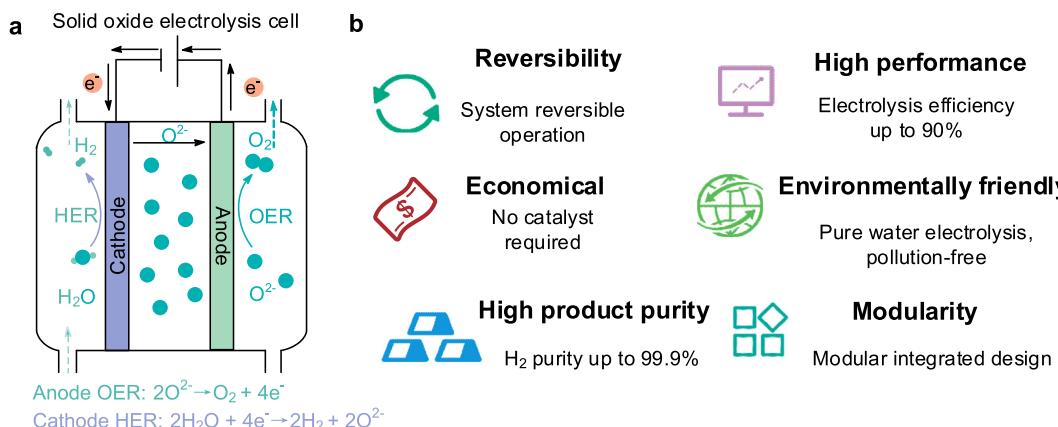


Fig. 1. Schematic diagram of the hydrogen production principle (a) and advantages (b) of SOEC.

of the most convenient and efficient tools to address those issues by providing insights into performance evolution and material degradation [11].

For example, Ni [12] developed a 2D thermal model to investigate the heat/mass transfer and chemical/electrochemical reactions in an SOEC and believed that reversible methanation and reforming reactions have a negative impact on the electrolysis performance of SOEC. Menon et al. [13] reported a numerical model to explain the functional interrelationship among various internal physico-chemical phenomena of an SOEC. To address the disadvantages of the above static models, Ali et al. [14] recommended that a dynamic model could provide more meaningful information for the improvement of SOEC. In this regard, Yin et al. [15] proposed a dynamic SOEC model to analyze the dynamic transients caused by various time constants of components. Zhang et al. [16] established and validated a real-time multiphysics model for pressurized SOEC. They suggested that SOEC should operate under the condition of the maximum endothermic point to maximize the use of thermal energy.

However, these models typically target a specific structure of SOEC and have poor generalization ability due to the use of a large number of physical parameters during the modeling process. In this regard, most of them are based on the existing material in the literature. Thus, these models fail to discover new SOEC materials and cannot reverse design the input parameters and features of SOEC according to actual needs. On the other hand, SOEC has been studied and developed for many years, accumulating a wealth of data and experience. Nevertheless, few reported works that utilize this data to construct SOEC models with good generalization ability.

To address these issues, machine learning (ML) can infer potential rules and relationships among material features, structural features, operational parameters, and output features of electrolytic cells with the aid of data collection [17]. For instance, Mohamed et al. [18] found that polynomial regression and logistic regression were the preferred algorithms to establish the ML model of PEM. Raeesi et al. [19] believed that the deep neural network algorithm had the highest prediction performance to analyze the performance of a PEM fuel cell stack. Kabir et al. [20] compared five ML models to improve the performance of green hydrogen production using dark fermentation and PEM. Bilgiç et al. [21] adopted an artificial neural network algorithm to predict the H₂ production of the magnetic field effect water electrolysis technology. Results indicate that the predicted results are in good agreement with the experimental data, promoting the optimal design of the electrolyzer. Li et al. [22] believed that ML has the potential to offer valuable insights for modeling and guiding biomass-to-syngas processes, transcending limitations associated with a single type or specific input of biomass gasification processes. Similarly, the ML-based approaches analyze the generalized structure-performance relationship of MOFs [23], photovoltaic-based [24], and electrochemical green hydrogen production [25].

These studies demonstrate that ML can be applied to predict and improve the performance of electrolysis technology by employing existent data. However, little literature has been reported on the application of ML to SOEC, let alone employ ML to reverse design and optimize SOEC. Therefore, a data-driven ML model of the SOEC system is established to determine its key parameters and explore their interaction relationships to achieve multi-objective optimization of the system. The main objectives of this study are to (i) develop and compare the random forest (RF) regression, support vector regression (SVR), deep neural network (DNN), and extreme gradient boosting (XGBoost) algorithms to select the core of ML for SOEC; (ii) identify and rank the key factors affecting electricity consumption and hydrogen production rate of SOEC using the Shapley additive explanations (SHAP) approach; (iii) investigate the effect of the key input features on the outputs of the SOEC system by conducting one-/two-/three partial dependencies plots (PDP) analyses; and (iv) balance the various output features of SOEC to achieve global optimization by coupling with the genetic algorithm (GA).

2. Methodology

The workflow of ML-assisted prediction and optimization of SOEC system is illustrated in Fig. 2. The literature data is collected and pre-processed to construct a dataset for the SOEC system. The dataset is used to train and test different ML models (including RF, SVR, DNN, and XGBoost models) to select the optimal model for the SOEC system after their hyperparameter tuning. The preferred ML model is integrated with the SHAP and partial dependencies plots analyses to identify the most important input features of the SOEC system and investigate their impacts on the target outputs. Finally, a genetic algorithm based multi-objective optimization is carried out to determine the optimal production conditions for the SOEC system.

2.1. Construction of experimental dataset

2.1.1. Data collection

A representative dataset is a prerequisite for the successful application of ML modeling. This study employed recognized literature search tools such as Google Scholar, X-mol, ScienceDirect, and Web of Science. A preliminary search using SOEC, high temperature steam electrolysis, and co-electrolysis of H₂O and CO₂ as the keywords yielded approximately 300 articles from 2012 to 2023. Based on the relevance of the papers, as well as the completeness and reliability of the data, approximately 50 articles focusing on electrolysis of H₂O or co-electrolysis of H₂O with CO₂ were ultimately selected to establish a dataset for ML models. All filtered data is impartially accepted without any preliminary assessment of its validity. For data not directly presented in tables or text, the open-source WebPlotDigitizer software is employed to extract the required information from graphs. Finally, a dataset of 355 data points was obtained as presented in Table S1. It encompasses 13 input features and 3 output features, as illustrated in Table 1. Among them, the input descriptors can be broadly classified into three categories: electrolytic cell descriptors, operational parameters, and feed conditions. Specifically, the electrolytic cell descriptors include the cathode electrode, surface area, electrolyte type, and electrolyte thickness. The operation parameters consist of humidity, voltage, flow, temperature, inlet gas composition, and pressure. The inlet conditions encompass the volume fractions of CO₂, H₂, and H₂O. The target features for ML models are chosen to be ohmic resistance, current density, and H₂ production rate in this study.

2.1.2. Correlation analysis of descriptors

To ensure the exclusion of relevant variables during the modeling process and enhance its complexity, it is imperative to assess the correlation among all descriptors employed in this study. The Pearson correlation coefficient between descriptors can be computed to accomplish this, as demonstrated in Eq. (1) [26]. Compared to other correlation analysis methods, the Pearson correlation coefficient has the following advantages [26]: (a) Simple and intuitive: the calculation method of the Pearson correlation coefficient is simple and clear, easy to understand and explain. It directly gives the degree of linear correlation between two variables. (b) Scope and explanation: the Pearson correlation coefficient ranges from -1 to 1, with -1 indicating complete negative correlation, 0 indicating no correlation, and 1 indicating complete positive correlation. This standardized range of values makes the interpretation of correlation coefficients more intuitive and unified. (c) Linear relationship detection: the Pearson correlation coefficient is mainly used to detect the linear relationship between two variables, so it is suitable for scenarios of linear correlation analysis. (d) Reliability and widely applicable: the Pearson correlation coefficient can provide reliable results and is widely used in research and practice in various fields, including social sciences, natural sciences, medicine, etc.

$$PCC = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{(X - \bar{X})^2 \sum (Y - \bar{Y})^2}} \quad (1)$$

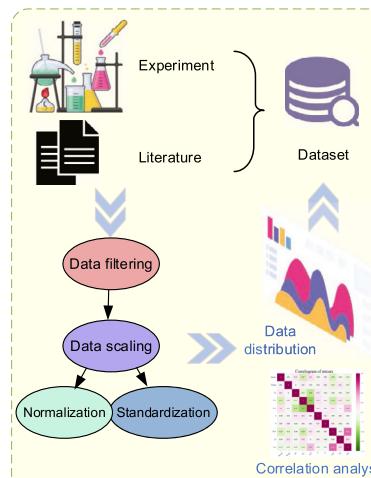
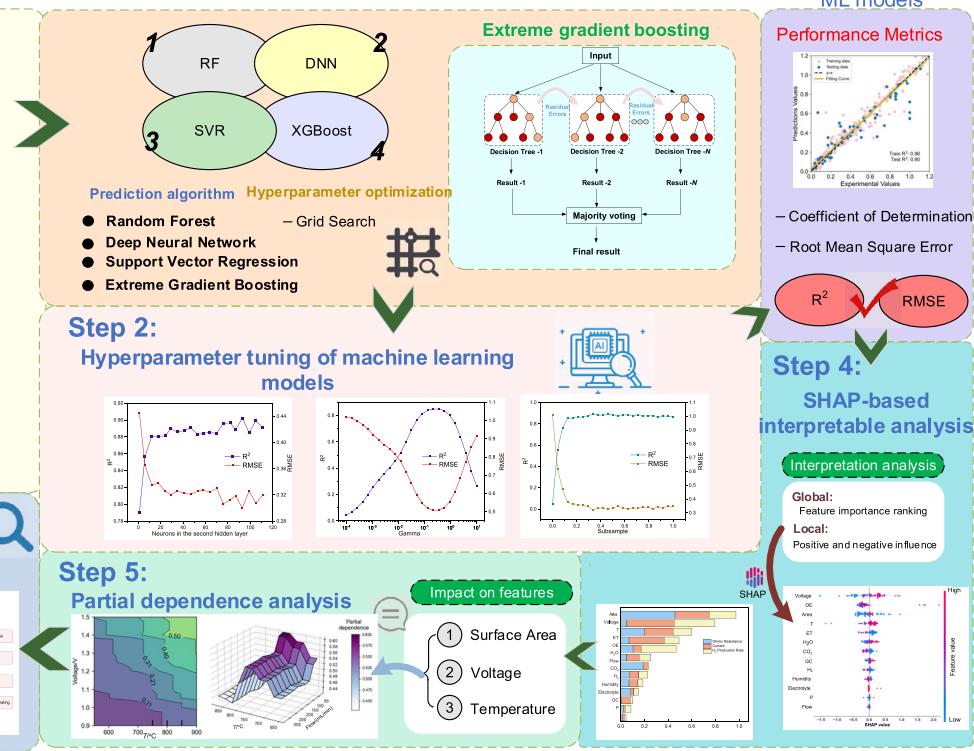
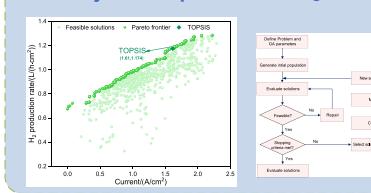
Step 1:**Construction of experimental dataset****Step 6:****Multi-objective optimization Algorithm**

Fig. 2. Schematic workflow of ML-assisted prediction and optimization of SOEC system.

Table 1

Input and output features of the ML models of the SOEC process.

Item	Variable type	Variable	Abbreviation	Unit
Input	Electrolytic cell (EC) descriptor	Cathode electrode	OE	/
		Surface area	Area	cm ²
		Electrolyte type	ET	/
		Electrolyte thickness		μm
Operation parameters (OP)	Humidity	/		%
		Voltage	/	V
		Volume flowrate	Flow	mL/min
		Temperature	T	°C
Feed conditions (FC)	Type of inlet gas composition	GC	/	
		Pressure	P	kPa
		Volume fraction of CO ₂	CO ₂	%
		Volume fraction of H ₂	H ₂	%
Output	Target features	Volume fraction of H ₂ O	H ₂ O	%
		Ohmic resistance	/	Ω/cm ²
		Current density	/	A/cm ²
		H ₂ production rate	/	L/(h·cm ²)

where PCC is the Pearson's correlation coefficient, X and \bar{X} are the sample and mean values of the x variable. Y and \bar{Y} are the sample and mean values of the y variable.

2.1.3. Data preprocessing

The process of feature scaling is a crucial part of data preprocessing, containing various techniques such as normalization, standardization,

and regularization. Data standardization can convert data into a distribution with a mean of 0 and a variance of 1, as depicted in Eq. (2). In comparison to alternative methods, it effectively mitigates the dimensional influence among different features, thereby promoting equitable weighting across features and enhancing both convergence speed and predictive performance of machine learning algorithms [27]. Hence, it is employed in this study. Additionally, dataset segmentation is another important preprocessing step in ML, serving to ensure the universality and accuracy of ML models through continuous testing and modification. In this regard, the raw data in the database is randomly shuffled with 80% of the data selected as the training set and the remaining 20% designated as the testing set. Simultaneously, a 5-fold cross-validation approach is used to cyclically utilize these subsets for training and validating models, thereby enhancing the evaluation of model performance.

$$x_i^* = \frac{x_i^n - \mu_i}{\sigma_i} \quad (2)$$

where x_i^* and x_i^n denote the standardized and true values of n^{th} sample feature belonging to i^{th} feature; μ_i refers to the average value of i^{th} feature; and σ_i is the standard deviation of the i^{th} feature.

2.2. Selection of the preferred ML models

2.2.1. Establishment of ML models

To identify the most suitable ML model for the SOEC process, this study evaluated and compared RF regression, SVR, DNN, and XGBoost as potential models as illustrated in Fig. 3a-d. Their basic principles are outlined as follows.

RF is a tree-based ensemble learning model with the bagging approach. Bagging employs the dropout sampling method on the original dataset to generate multiple training sets, which are then used to train individual base learners. In this process, parallel training can be

conducted due to the independent nature of each base learner, followed by combining the trained learners. Multiple subtrees are constructed through a random selection of features and samples, and these subtrees collectively contribute to the final prediction result through voting. In this study, the RandomForestRegression class from the Scikit Learn library is employed for implementing RF regression. The SVR method is widely employed in regression analysis, wherein it projects data onto a high-dimensional space to identify a hyperplane that minimizes sample errors within a specified range. The DNN algorithm is a computational model that simulates the biological nervous system, consisting of multiple neurons, each with an activation function, and capable of receiving input from other neurons. The XGBoost algorithm is a ML algorithm based on gradient boosting trees to enhance the performance of weak base learners through stacking. Compared with other gradient boosting methods, the XGBoost algorithm adopts an improved regularization technique to avoid overfitting and improve the model's generalization ability.

2.2.2. Hyperparameter tuning of ML models

The objective of hyperparameter tuning is to identify the optimal combination of hyperparameters in ML models, aiming to enhance model performance and generalization capability, minimize bias and variance,

and thereby improve predictive accuracy on new data. Adjusting hyperparameters such as model complexity, regularization degree, and learning rate, can influence the training process and overall model performance. When dealing with a wide parameter range in ML models, the grid search approach is particularly suitable as it enables comprehensive exploration of various hyperparameter combinations leading to improved model performance compared to other approaches [28]. Therefore, the grid search approach is selected in this work. Specifically, this study initially conducts an extensive search and subsequently refines the optimal hyperparameters within a narrower range. This approach effectively reduces the search space for hyperparameters, enabling the identification of the optimal combination within a more constrained scope. Moreover, this methodology minimizes computational resource consumption while achieving closer proximity to the optimal solution within a given timeframe. The detailed range of the hyperparameter of each ML model is illustrated in Table 2.

2.2.3. Evaluation of ML models

After conducting hyperparameter optimization and training, the prediction accuracy of four ML models was assessed on the testing set. This evaluation aids in identifying the optimal ML model for the SOEC process and facilitates subsequent feature analysis. This paper employs

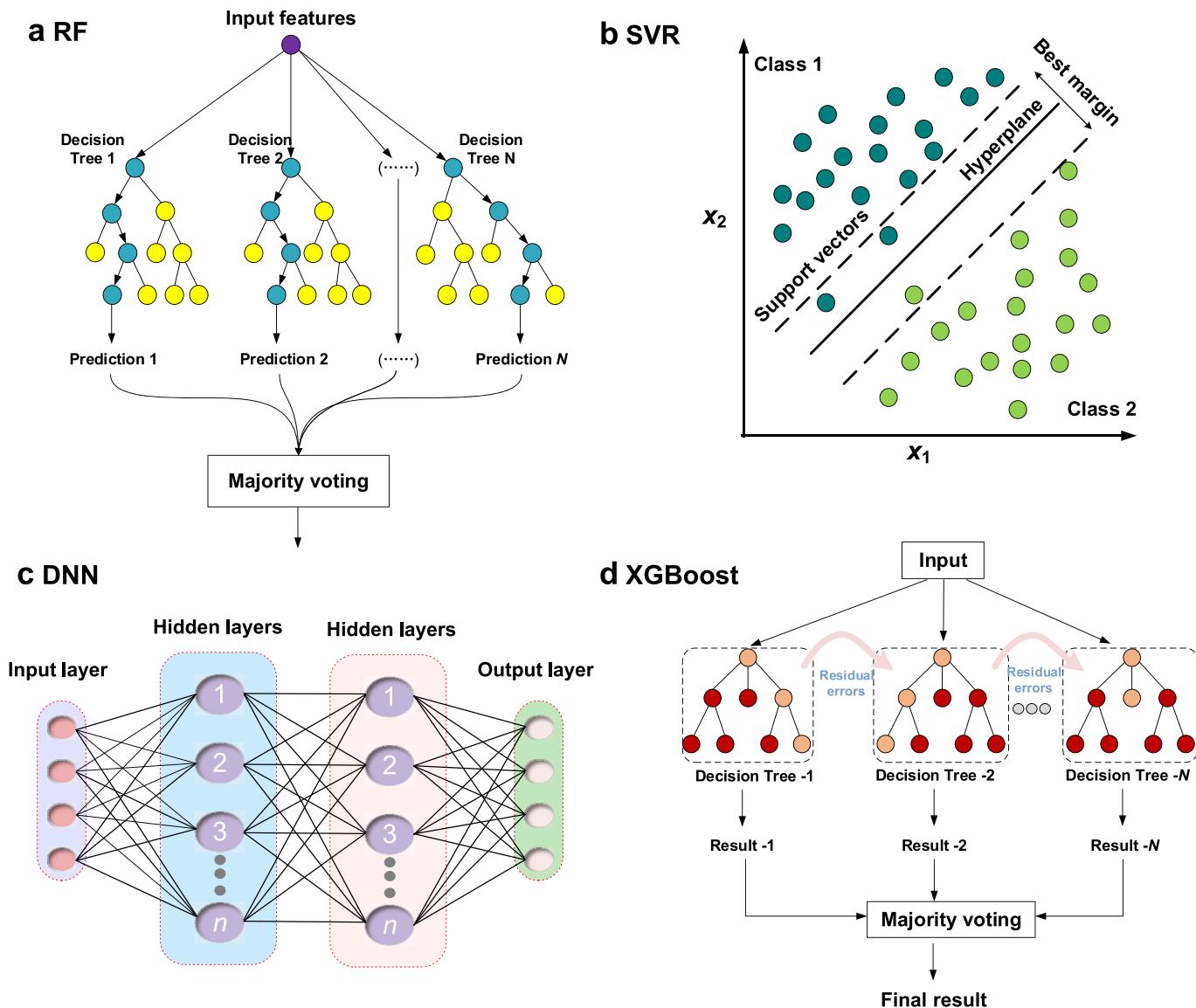


Fig. 3. Schematic diagram of four ML models: (a) RF, (b) SVR, (c) DNN, and (d) XGBoost.

the coefficient of determination (R^2) and root mean square error (RMSE) to assess the optimization outcomes of the ML models with various learning algorithms based on actual and predicted output target values as formulated in Eqs. (3) and (4) [33]. The R^2 approach is to first calculate the square difference and total square difference between the predicted values of the model and the actual observed values. By dividing the squared difference by the total squared difference, a ratio ranging from 0 to 1 is obtained. Subsequently, subtracting this proportion from 1 and taking its negative value yields R^2 . A higher R^2 indicates a stronger explanatory power of the model for the target variable, implying better data fit. The RMSE quantifies the disparity between the predicted values of a model and the actual observed values. A smaller RMSE value indicates a diminished discrepancy between the predicted and observed values, thereby signifying enhanced predictive capability of the model. Therefore, a higher R^2 value and a lower RMSE value are indicative of superior model fit and enhanced predictive performance.

$$R^2 = 1 - \frac{\sum_{k=1}^n (y_k - \hat{y}_k)^2}{\sum_{k=1}^n (\hat{y}_k - \text{mean}(y))^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2} \quad (4)$$

Where y_k and \hat{y}_k represent the experimental and predicted values; $\text{mean}(y)$ is the average of the corresponding experimental value; n and k denote the total number of samples and the sample index, respectively.

2.3. Interpretable analysis of ML model

ML models are often perceived as opaque due to the challenge users face in comprehending the underlying mechanisms of these models, particularly when analyzing how input features are mapped to output results. Consequently, it becomes arduous to elucidate the relationship between material properties and performance using ML models. This study introduces the SHAP method for interpretive analysis of such models. SHAP is a game theory-based approach that offers intuitive explanations of feature variable importance, enhancing the interpretability of ML models, as demonstrated in Eq. (5) [34]. In particular, it can not only demonstrate the degree of influence of features on model output, but also clearly display their positive and negative effects.

$$f(x_i) = f_{\text{base}} + \sum_{j=1}^n \phi(x_{i,j}) \quad (5)$$

Where x_i represent the i^{th} sample of the model; $f(x_i)$ represents the predicted value of the model for the i^{th} sample; f_{base} refers to the baseline of the model, which is the average value of the dependent variable for all samples; $\phi(x_{i,j})$ represents the contribution of the j^{th} feature in the i^{th} sample to the predicted value.

Table 2
Detailed hyperparameters of each ML model for hyperparameter tuning.

Algorithm	Hyperparameter	Range [29–32]
RF	n_estimators	10–200
	max_depth	5–30
	min_samples_split	2–10
	min_samples_leaf	1–5
SVR	gamma	0.001–10
	C	0.1–100
	kernel	linear, poly, rbf
DNN	Neurons in the first hidden layer	10–200
	Neurons in the second hidden layer	10–200
	max_iter	100–8000
XGBoost	learning_rate	0.01–1
	n_estimators	10–200
	subsample	0.1–1

2.4. Sensitivity analysis

Partial Dependence Plots (PDPs) are a valuable tool for visualizing the relationship between input variables and outputs in ML models. By plotting the average or expected value of the output against one or more input variables, partial dependence curves can demonstrate the overall impact of each variable on the output and reveal nonlinear relationships and interactions. Therefore, the PDP analysis is carried out to investigate the impact of input characteristic variables on the current density and hydrogen production of the SOEC process, which can enhance comprehension regarding the relationship between model input variables and output, as well as provide a theoretical foundation and technical support for determining optimal operating conditions in the SOEC process.

2.5. Multi-objective optimization of SOEC

The expected SOEC is to produce more hydrogen gas at lower power consumption. Therefore, this work aims to minimize the current density and maximize the hydrogen production rate as objective functions as indicated in Eq. (6). Considering the significant influence of electrolytic cell descriptors, operation parameters, and feed conditions on SOEC performance, the upper and lower limits of these parameters are selected as constraints for multi-objective optimization in Eqs. (7)–(9). By integrating the genetic algorithm (GA) with the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method, the optimal parameter combination is determined to achieve an ideal trade-off between the output current density and hydrogen production rate.

$$\text{Objective functions : } \max F = \omega_1 f_{\text{H}_2}(\text{EC}, \text{OP}, \text{FC}) - \omega_2 f_{\text{i}}(\text{EC}, \text{OP}, \text{FC}) \quad (6)$$

Constraints:

$$\text{LB of EC} \leq \text{OE, Area, Electrolyte, CT, ET} \leq \text{UB of EC} \quad (7)$$

$$\text{LB of OP} \leq \text{Humidity, Voltage, Flow, T, GC, P} \leq \text{UB of OP} \quad (8)$$

$$\text{LB of FC} \leq \text{H}_2, \text{CO}_2, \text{H}_2\text{O} \leq \text{UB of FC} \quad (9)$$

where LB and UB refer to the lower and upper bounds; f_{H_2} and f_{i} denote the function of the hydrogen production rate and current density of SOEC; and ω_1 and ω_2 are the weight factor of these two parameters, which are same to 0.5.

3. Results and discussions

3.1. Dataset visualization analysis

3.1.1. Input feature data distribution

The data distribution can facilitate a comprehensive understanding of the overall status of the established SOEC dataset, as well as provide insights into the intervals representing various input features. The distribution of the dataset for the SOEC process is illustrated in Fig. 4a–i. The triangle and horizontal lines within the rectangle represent the mean and mode values of each feature. The modes and averages of the nine input features are very close, especially in terms of voltage, temperature, and the volume fractions of CO_2 and H_2O , which are basically equal to 1.2 V, 800 °C, 0.25, 0.45, as shown in Fig. 4b and d–f. The modes of the humidity and volume fraction of H_2 are exhibited at the upper boundary of the rectangle (40 and 0.2), slightly surpassing their respective mean values (38.0 and 0.17) as depicted in Fig. 4a and c. Similarly, the mode of electrolyte thickness is higher than its average, with values of 10 μm and 18.37 μm , respectively. On the contrary, the modes of volume flowrate and surface area (100 mL/min and 1 cm^2) are slightly smaller than their respective averages (125 mL/min and 1.35 cm^2) as indicated in Fig. 4g and i. According to the height of the data distribution rectangle of the nine input features, it can be seen that the nine input features cover a

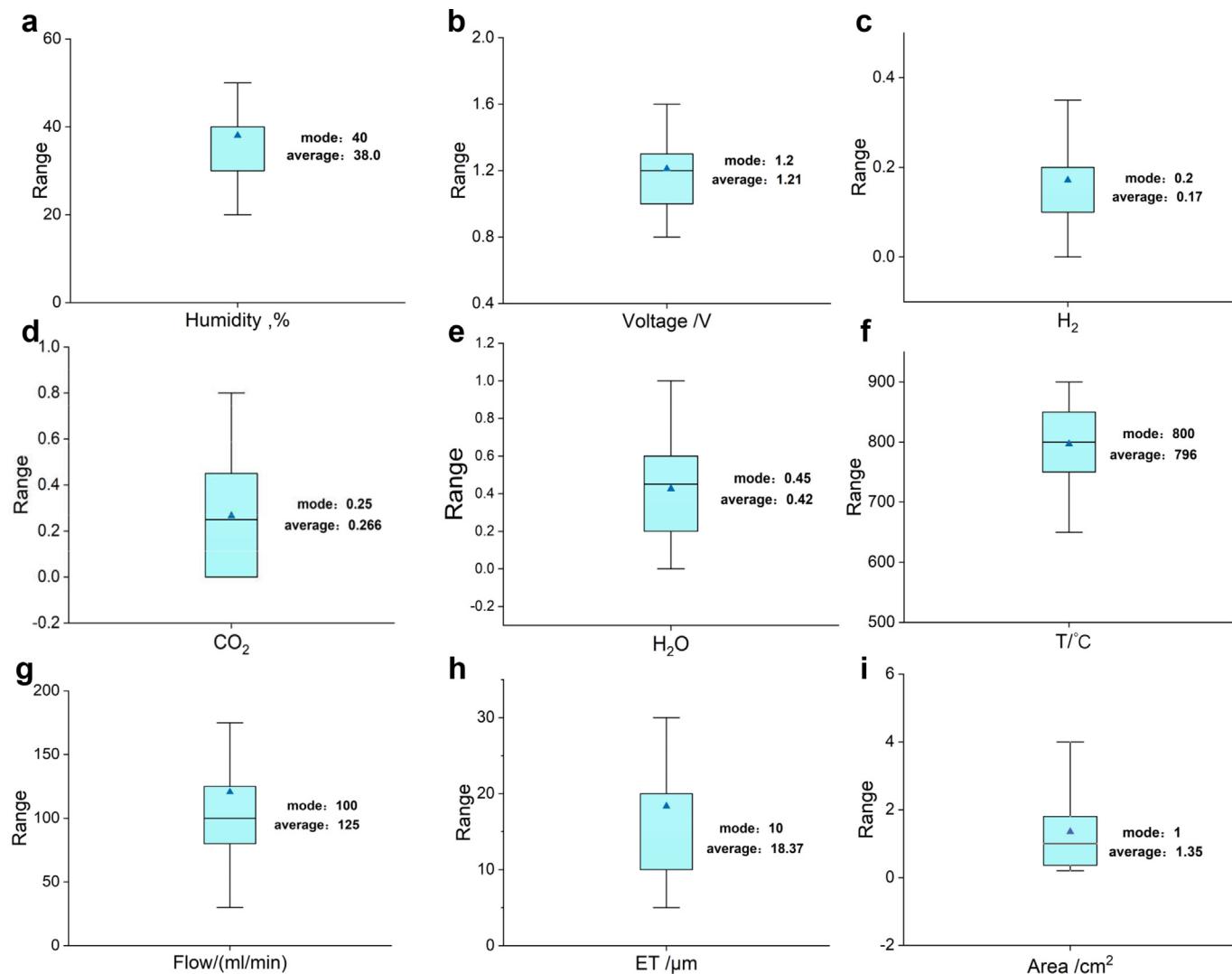


Fig. 4. Data distribution, mode, and average of the input features of the SOEC process.

large input range. Therefore, the established dataset has good representativeness and can be used for training and testing the following different machine learning models.

3.1.2. Descriptive analysis

The Pearson correlation coefficient (PCC) is employed to quantify the linear correlation between two input variables, thereby eliminating redundancy from the obtained feature subset and facilitating the efficient construction of ML models. The Pearson correlation matrix of the SOEC process is shown in Fig. 5. The red and green tones in the figure represent positive and negative correlations, while the magnitude of their values signifies the strength of correlation between the two features. A higher absolute value indicates a stronger linear relationship between variables. The results show that the correlation coefficients of the vast majority of input features are less than 0.5, except for the notable correlation observed between flow and effective area, which demonstrates a maximum coefficient of 0.63. Therefore, the selected input features in this study have no strong linear correlation and can serve as input descriptors for the establishment of machine learning models of the SOEC process.

3.2. Optimization of the hyperparameters in ML models

Four representative ML models, RF, DNN, SVR, and XGBoost models,

for the SOEC process were constructed using the dataset collected and established in Section 2.1. To ensure optimal predictive performance of these machine learning models, this study first optimized the hyperparameters of these models and evaluated them by R^2 and RMSE, as illustrated in Figs. 6a-e, S2, and S3.

The results depicted in Fig. 6a demonstrate that as the penalty coefficient increases from 0.1 to 74, the R^2 of the SVR model rapidly increases while the RMSE significantly decreases. However, when the penalty coefficient exceeds 74, no noteworthy changes are observed in both R^2 and RMSE. Consequently, it is recommended to set the penalty coefficient for the SVR model at 74. As the gamma of the SVR model increases from 10^{-4} to 0.74 (Fig. 6b), there is a significant improvement in the R^2 value and a rapid decrease in RMSE. However, exceeding a gamma value of 0.74 leads to a notable decline in the predictive performance of the model. Therefore, the optimal gamma of the SVR model is set to 0.74 in this work.

The tuning of the XGBoost algorithm with different hyperparameters is illustrated in Fig. 6c-e. When the number of estimators ($n_{\text{estimators}}$) is greater than 50, both the R^2 and RMSE values of the XGBoost model exhibit negligible changes, indicating a convergence toward stable predictive performance as shown in Fig. 6c. Similarly, the XGBoost model exhibited good predictive performance at learning_rate and subsample of 0.12 and 0.45, as illustrated in Fig. 6d and e. Adding these two hyperparameters further does not significantly improve the performance of the

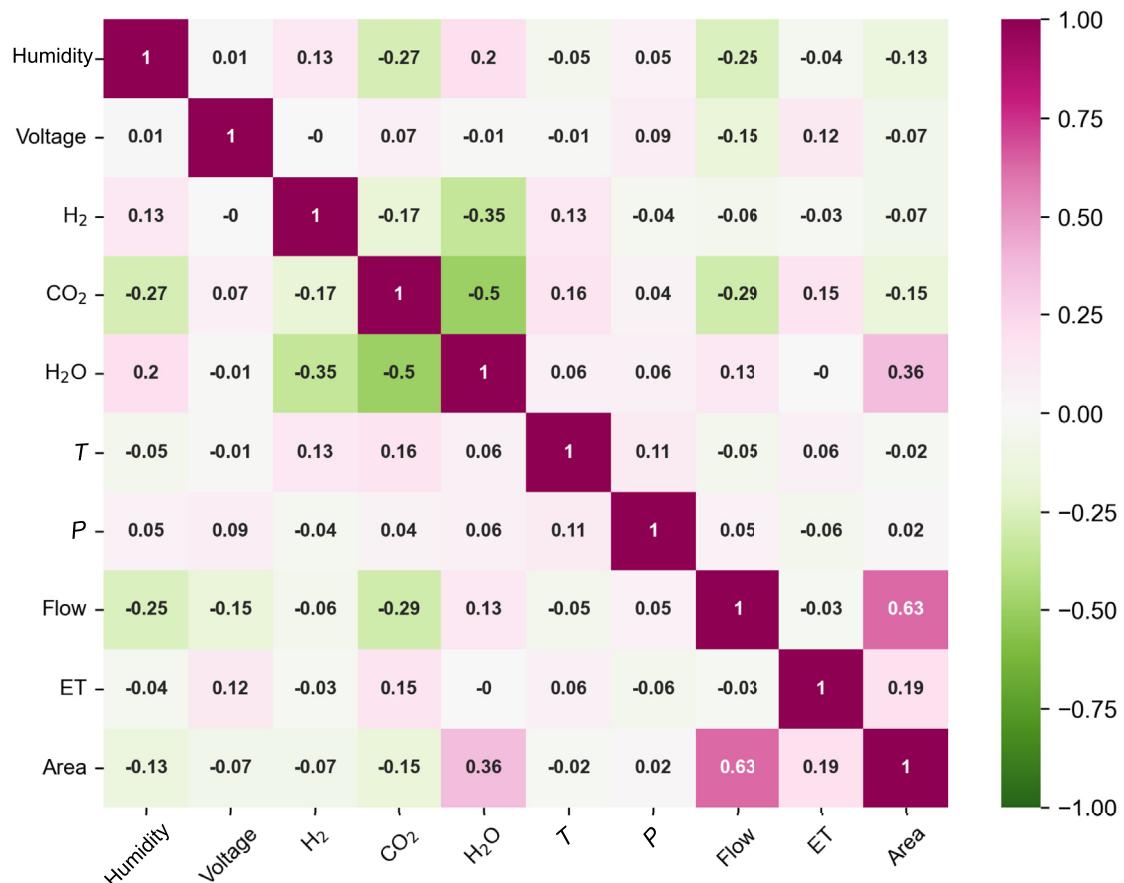


Fig. 5. Pearson heatmap of input features for the modeling of the SOEC process.

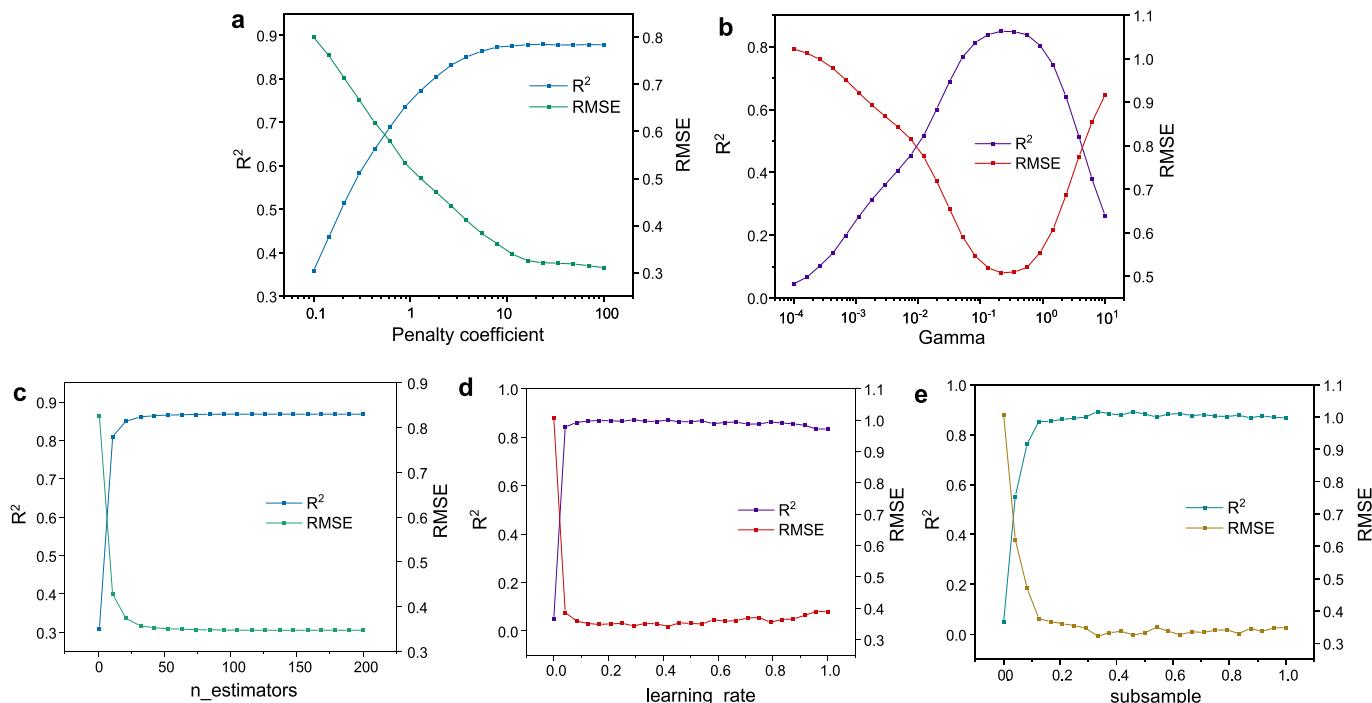


Fig. 6. Optimization of the hyperparameters of the SVR (a,b) and XGBoost (c–e) models (Those of the RF and DNN models are indicated in Figs. S2 and S3).

model. Hence, the n_estimators, learning_rate, and subsample of the XGBoost model are set to 55, 0.12, and 0.45, respectively. In addition, the main hyperparameters of the RF and DNN models are also optimized as shown in Figs. S2 and S3. The R² value of the RF model exhibits a significant increase as the number of estimators or maximum depth (max_depth) increases as shown in Figs. S2 and S3. The RF model attains its optimal performance when n_estimators and max_depth are set to 104 and 27, respectively. The optimization of DNN primarily focuses on analyzing the impact of the number of neurons in the first and second hidden layers on the predictive performance of the model as illustrated in Fig. S3a and 3b. The results reveal that when the model reaches its maximum R², the corresponding numbers of neurons in the first and second hidden layers are 104 and 92, respectively. In summary, the results of optimizing the key hyperparameters of the four ML models of the SOEC process are summarized in Table 3.

3.3. Comparison of the ML models of the SOEC process

After optimizing the key hyperparameters of RF, DNN, SVR, and XGBoost models, 20% of the total data was randomly selected as the testing set, while the remaining 80% constituted the training set (ensuring their mutual exclusivity). Subsequently, the R² and RMSE of four different ML models were compared to select the optimal ML for the SOEC process after 5-fold cross-validation as indicated in Fig. 7a-l. Comparing Fig. 7a-d, it can be observed that the predictive performance of RF, DNN, and XGBoost models for Ohmic resistance is roughly equivalent, as indicated by their test set R² values ranging from approximately 0.94 to 0.96 except that of the SVR model is only 0.81. Regarding the prediction of current intensity, results reveal that the fitting degrees of RF, DNN, SVR, and XGBoost models to current intensity are 0.93, 0.90, 0.84, and 0.95, respectively, as shown in Fig. 7e-h. It indicates that the predictive performance of XGBoost model surpasses that of the other three algorithms. This advantage is further demonstrated by accurately predicting the rate of hydrogen production as illustrated in Fig. 7i-l, it can be observed that the R² of RF, DNN, SVR, and XGBoost models for hydrogen production rate are 0.92, 0.92, 0.81, and 0.97, respectively. More information on the prediction results of RF, DNN, SVR, and XGBoost models for Ohmic resistance, current intensity, and hydrogen production rate (including the training set and test set) is summarized in Table 4. Therefore, considering the comprehensive evaluation of R² and RMSE for each ML model's three output targets, it can be found that the XGBoost algorithm has the best predictive performance. Consequently, it is selected for subsequent analysis and optimization of catalyst performance.

3.4. SHAP-based interpretable analysis of the preferred ML model

The SHAP method is employed to analyze the impact of each input feature of the selected XGBoost model on the output parameters to enhance the understanding of the predictive behavior of the model as demonstrated in Figs. 8, 9, S4, and S5. The global feature importance

ranking based on SHAP values is illustrated in Fig. 8. Compared to operational parameters and feed conditions, the descriptor of the electrolytic cell is more important for the output parameters of the ML model of SOEC, accounting for 54.5%. In particular, the effective area, voltage, and temperature of the SOEC have the greatest impact on the three output results, accounting for 21.6%, 16.6%, and 13.0%, respectively.

In addition, feature importance analysis based on the SHAP method also has local interpretability. For example, the specific impact results of the contribution and importance ranking of each input feature on the three output indicators are depicted in Fig. 9a and b, Figs. S4, and S5. Fig. 9a and b demonstrate voltage is particularly crucial for the H₂ production rate. The SHAP values on the right indicate that as voltage increases, its influence on the H₂ production rate gradually intensifies. Particularly at low voltages, voltage may have a negative impact on the H₂ generation rate; however, as voltage rises, its effect shifts towards positive. Regarding to the Ohmic resistors, the effective area, temperature, and feed flow rate exert the most significant influence on them as shown in Fig. S4. Furthermore, the voltage also significantly influences the current density of SOEC as indicated in Fig. S5. Additionally, the current density decreases with increasing electrolyte thickness (color changes from red to blue), which is consistent with the results in the literature [35]. It is observed that electrolyte plays a crucial role in ion transport within the SOEC. With thicker electrolyte layers, ions have longer transport paths leading to increased resistance and subsequently reduced current density at identical voltages, ultimately minimizing electrical energy consumption.

3.5. Partial dependence analysis of the key parameters of SOEC

The SHAP values effectively identify the key input features influencing the three objective variables of the SOEC process. The impact of these crucial input features on their respective output variables is analyzed using one-dimensional, two-dimensional, and three-dimensional partial dependency plots. Due to the two- and three-dimensional partial dependency analyses can reflect more information compared to that of one-dimensional partial dependency, they are mainly analyzed and explained in the subsequent sections, while the results of the one-dimensional analysis are predominantly presented in Fig. S6.

3.5.1. Two-dimensional partial dependence analysis

The results of the partial dependence analysis for the three key input characteristics of Ohmic resistance, namely effective area, temperature, and volumetric flow rate, are depicted in Fig. 10a-c. At lower gas flow rates (< 50 mL/min) and lower temperatures (< 800 °C), the Ohmic resistance can be significantly increased as shown in Fig. 10a. This is mainly due to the limitation of interface reaction rate under low temperature and low flow rate conditions, leading to an increase in electron transfer resistance between the electrolyte and electrode, thereby increasing Ohmic resistance. Fig. 10b shows that the Ohmic resistance of SOEC is relatively high at higher temperatures due to potential issues faced by solid oxide electrolytes and electrode materials in such environments, including thermal expansion, changes in crystal structure, or chemical stability. These factors can negatively impact ion and electron transport, leading to an increase in Ohmic resistance. However, it is evident that the effective area plays a more significant role in determining the Ohmic resistance of SOEC as it exhibits a monotonic decreasing trend with increasing effective area. According to the results depicted in Fig. 10c, the influence of effective area variation (0.2–2 cm²) on Ohmic resistance remains insignificant at lower gas flow rates (< 50 mL/min). This is attributed to the fact that within this range, the gas flow rate fails to generate a substantial impact on Ohmic resistance due to alterations in the effective area.

The current density is significantly influenced by three characteristics, namely voltage, electrolyte thickness, and effective area, as demonstrated in Fig. 11a-c. When the voltage exceeds 1.3 V, the current density correspondingly increases. Simultaneously, while electrolyte

Table 3
Optimized hyperparameters of different ML models of the SOEC process.

Algorithm	Hyperparameter	Optimized value
RF	n_estimators	60
	max_depth	27
	min_samples_leaf	2
SVR	gamma	0.74
	C	74
	kernel	rbf
DNN	Neurons in the first hidden layer	104
	Neurons in the second hidden layer	92
	max_iter	1000
XGBoost	learning_rate	0.12
	n_estimators	55
	subsample	0.45

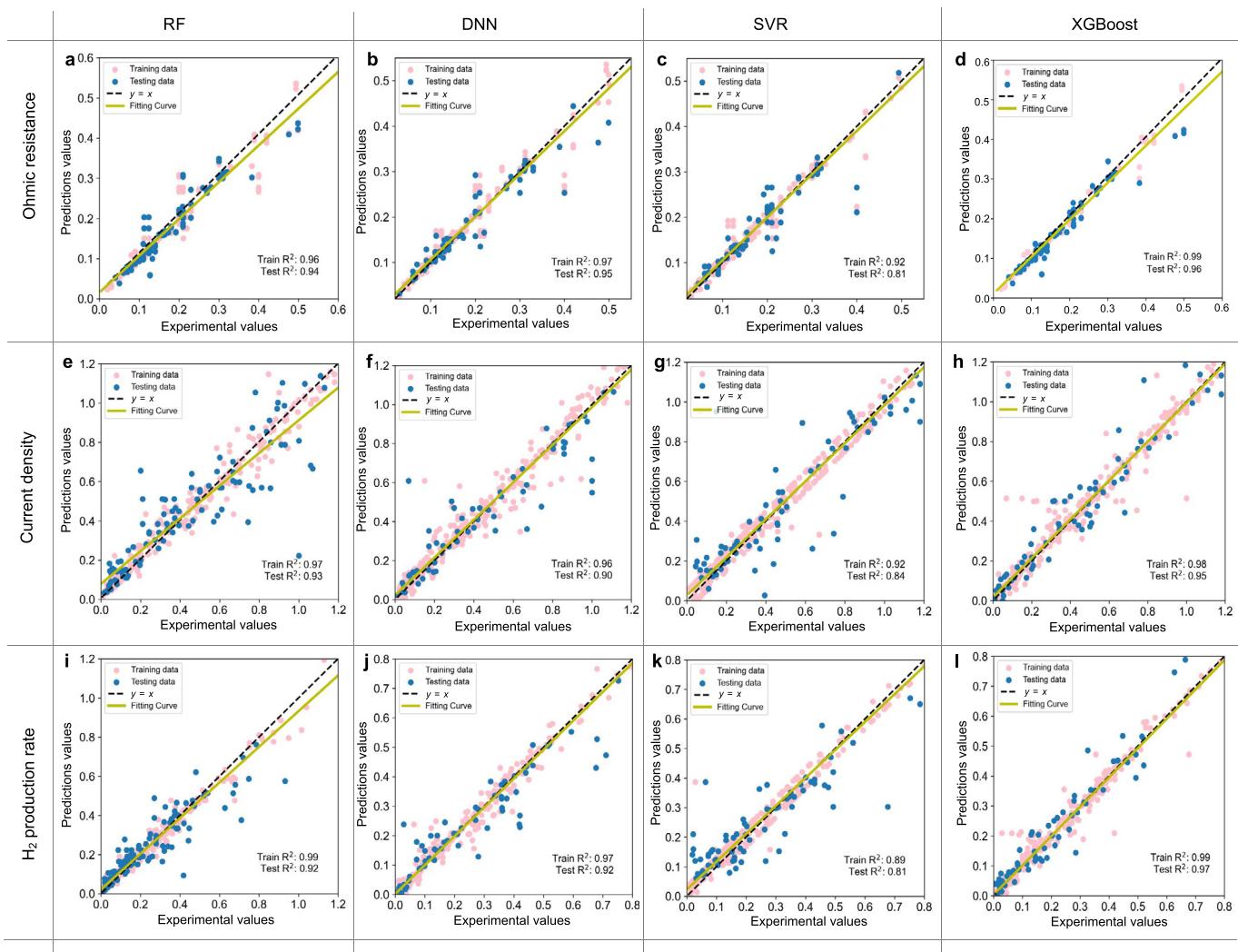


Fig. 7. Comparison of the predictive performance of four ML models: Ohmic resistance (a-d), current density (e-f), and H₂ production rate (g-l).

Table 4
Comparison of the R² and RMSE of the predicted performance of the SOEC's ML models.

Models	Item	Ohmic resistance	Current density	H ₂ production ratio
RF	Train R ²	0.9563	0.9694	0.9855
	Train	0.0258	0.0812	0.0352
	RMSE			
	Test R ²	0.9450	0.8349	0.9180
	Test RMSE	0.0325	0.2255	0.1014
	Train R ²	0.9728	0.9628	0.9716
DNN	Train	0.0213	0.0723	0.0293
	RMSE			
	Test R ²	0.9471	0.8993	0.9216
	Test RMSE	0.0346	0.1800	0.0826
	Train R ²	0.9215	0.9192	0.8944
	Train	0.0765	0.0762	0.1105
SVR	RMSE			
	Test R ²	0.8143	0.8406	0.8062
	Test RMSE	0.1020	0.1022	0.2241
	Train R ²	0.9794	0.9792	0.9886
	Train	0.0210	0.0704	0.0347
	RMSE			
XGBoost	Test R ²	0.9620	0.9536	0.9677
	Test RMSE	0.0316	0.1664	0.0817
	Train R ²	0.9794	0.9792	0.9886
	Train	0.0210	0.0704	0.0347

thickness has a certain impact on current density, its influence is not as significant as that of voltage. In cases where the effective area of SOEC is less than 0.4 cm² and the voltage surpasses 1.3 V, the current density can reach its maximum value due to heightened reaction rates resulting from

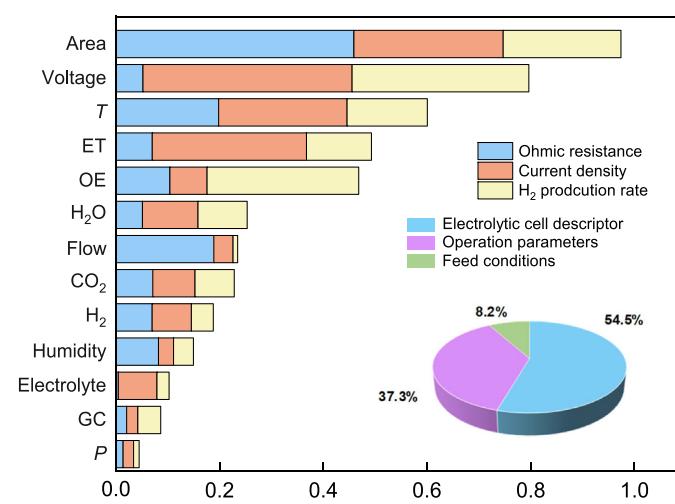


Fig. 8. SHAP-based global interpretation of the XGBoost model.

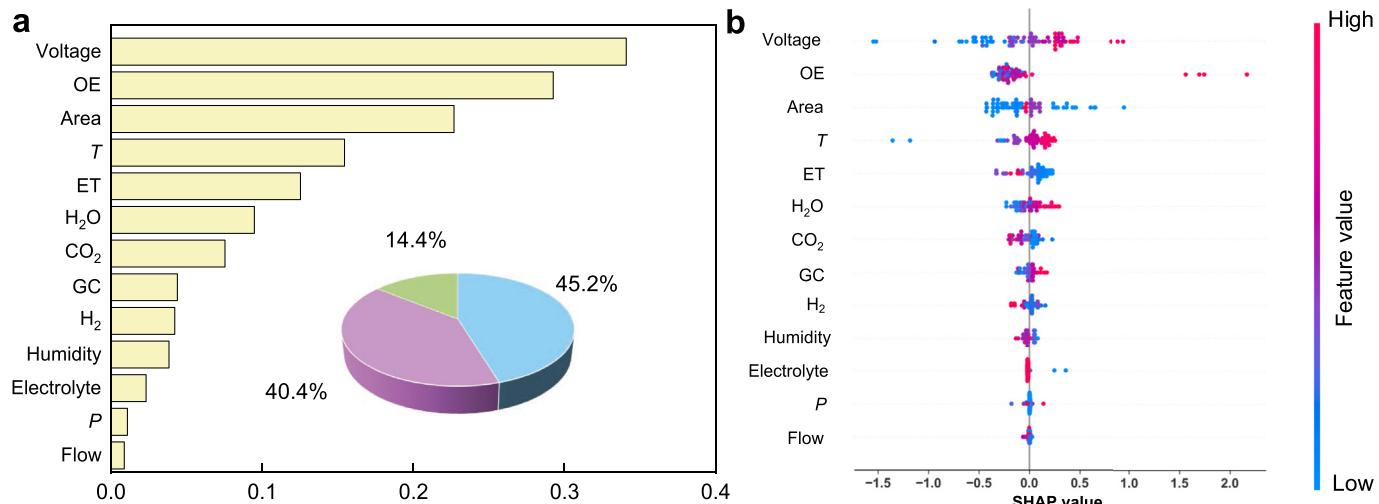


Fig. 9. Local interpretation of H₂ production rate (a and b). (That of the Ohmic resistance and current density is shown in Figs. S4 and S5).

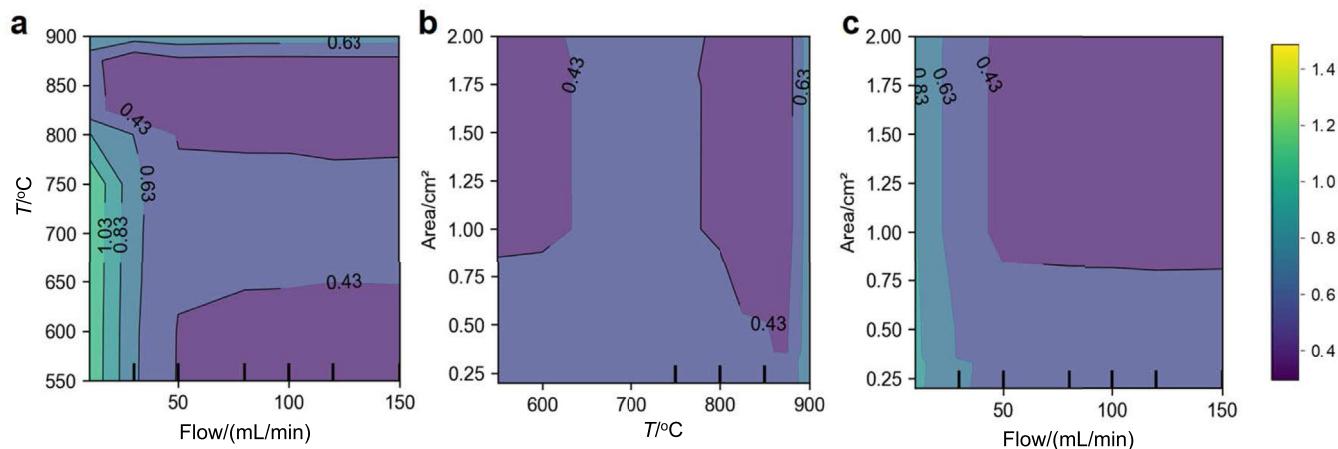


Fig. 10. Two-dimensional partial dependence plots of Ohmic resistance influenced by the corresponding most important input variables.

high voltage levels. The electrolyte thickness affects ion diffusion rates and indirectly impacts current density. Therefore, the performance of SOECs improves with decreasing electrolyte thickness.

In terms of predicting the rate of H₂ production, this study focuses on the effects of voltage, effective area, and temperature, as shown in Fig. 12a–c. According to the results depicted in Fig. 12a, the study reveals that the rate of H₂ production exhibits a negative correlation with the effective area of SOEC while showing a positive correlation with temperature. It is mainly because an increase in the effective area of SOEC leads to a decrease in current density on the electrode, resulting in a reduction in the water electrolysis rate. Consequently, as the effective area of the SOEC increases, there is a corresponding decrease in the H₂ gas production rate; however, no significant change is observed when the effective area exceeds 0.5 cm². Under constant effective area conditions, the H₂ production rate demonstrates an upward trend with increasing temperature. This can be explained by enhanced molecular kinetic energy at higher temperatures, facilitating easier passage through mass transfer interfaces and overcoming activation energy barriers for reaction initiation, thereby accelerating H₂ production, and improving its overall rate. According to the temperature-voltage interaction diagram depicted in Fig. 12b, it indicates that the SOEC has an optimal H₂ production rate under high-temperature conditions (800–900 °C) and elevated voltages (> 1.3 V). It is due to the comprehensive utilization of SOEC's electrolysis

performance at higher temperatures and voltages, thereby facilitating an enhanced H₂ production rate. Consequently, operating SOEC at higher temperature and voltage levels enables more efficient electrolysis reactions and yields superior rates of H₂ generation. These findings also offer valuable insights for optimizing the operational parameters of SOEC. The interaction diagram depicted in Fig. 12c illustrates that a decrease in the effective area leads to an increase in voltage, consequently resulting in an enhanced rate of hydrogen production. This is because the decrease of the effective area could subsequently reduce the contact area between the electrode and electrolyte, leading to an increase in the current density per unit area and the efficiency of the electrolysis reaction.

3.5.2. Three-dimensional partial dependence analysis

The three-dimensional partial dependence of the most important input features on three outputs is analyzed by the preferred XGBoost model and illustrated in Figs. 13–15. The Ohmic resistance gradually decreases as the effective area of SOEC changes from 0 to 4 cm², as depicted in Fig. 13a. However, beyond an effective area of 4 cm², the Ohmic resistance exhibits a gradual increase until it reaches a plateau after surpassing an effective area of 20 cm². This is attributed to the positive correlation between electrolysis temperature and electrolyte solubility, resulting in an elevated concentration of the electrolyte. The

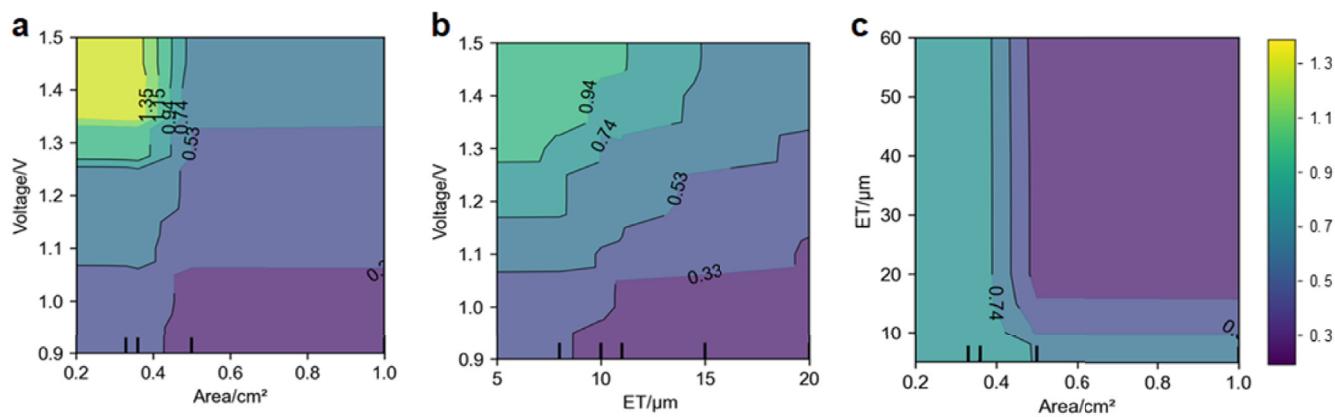


Fig. 11. Two-dimensional partial dependence plots of current density influenced by the corresponding most important input variables.

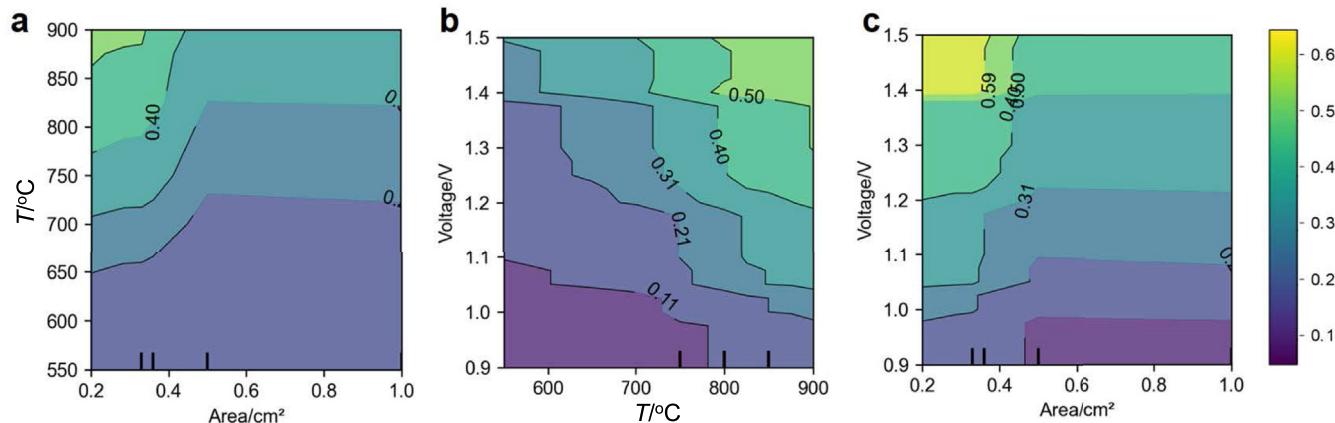


Fig. 12. Two-dimensional partial dependence plots of H_2 production rate influenced by the corresponding most important input variables.

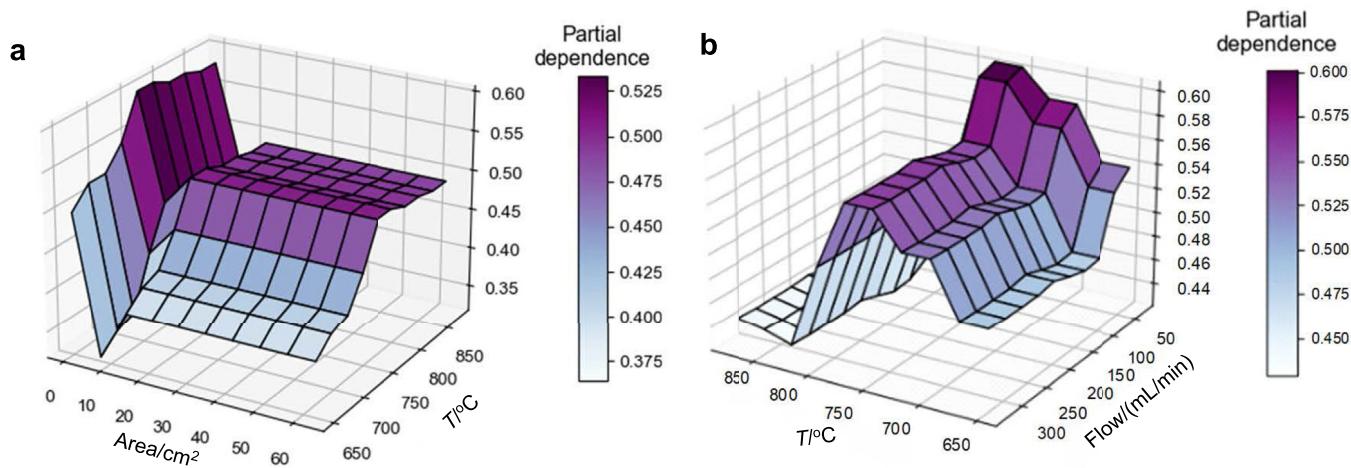


Fig. 13. Three-dimensional PDP of the impact of the most important input variables on Ohmic resistance.

increased concentration of the electrolyte subsequently reduces its resistivity, facilitating efficient electrical conduction and diminishing overall Ohmic resistance for enhanced hydrogen gas production through easier electrolysis. Consequently, high temperatures (> 800 °C) and relatively low specific surface areas ($4-10\text{ cm}^2$) create favorable conditions for electrolysis. According to Fig. 13b, the Ohmic resistance reaches its maximum value when the gas flow rate is below 50 mL/min. As the

gas flow rate increases to 100 mL/min, there is no further change in the Ohmic resistance. However, the Ohmic resistance of SOEC is significantly decreased if the temperature is higher than 800 °C. Therefore, for achieving higher electrolysis efficiency and facilitating hydrogen gas production, it is recommended to operate at temperatures above 800 °C and with a gas flow rate exceeding 50 mL/min.

The three-dimensional partial dependence of current density on

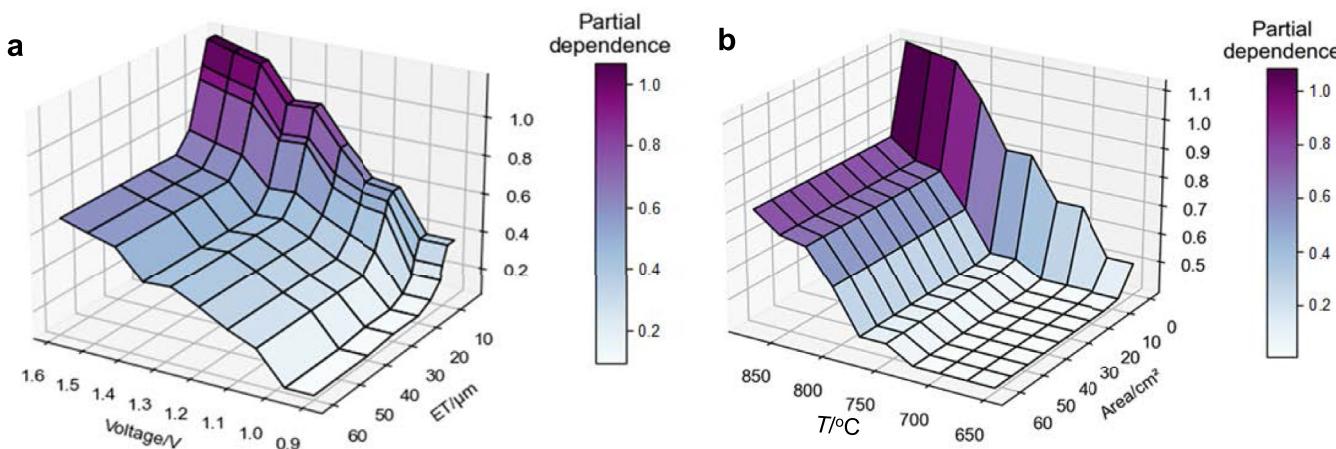


Fig. 14. Three-dimensional PDP of the impact of the most important input variables on current density of SOEC.

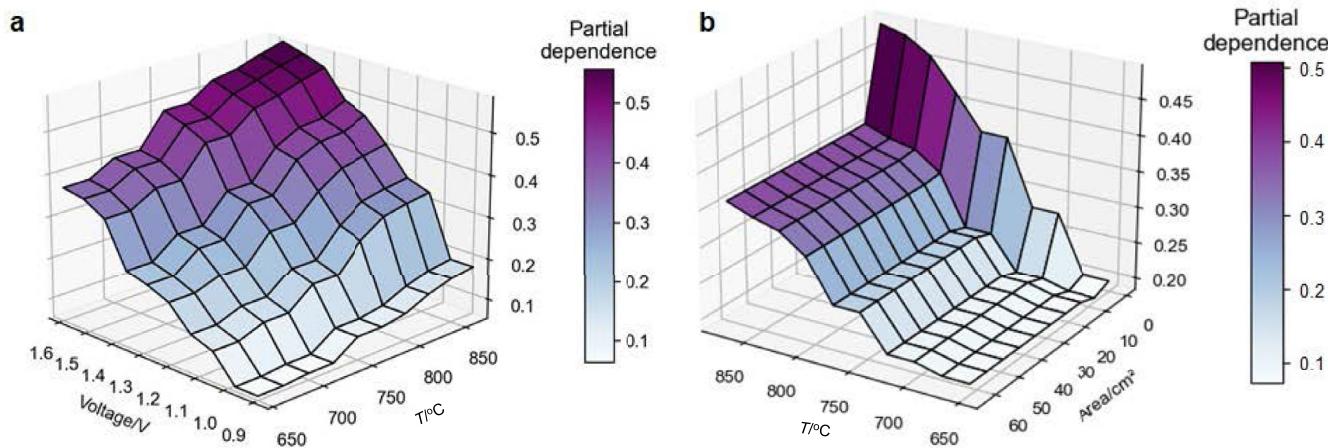


Fig. 15. Three-dimensional PDP of the impact of the most important input variables on the rate of H₂ production of SOEC.

voltage, temperature, and electrolyte thickness (ET) is illustrated in Fig. 14a and b. Fig. 14a reveals that the current density reaches its maximum value under low thickness (< 20 μm) and high voltage (> 1.4 V) conditions. Within a certain range, increasing the current density enhances the rate of H₂ production. However, when the thickness of the electrolyte exceeds 20 μm, there is a sharp decline in current density as observed in Fig. 14a. Furthermore, Fig. 14b demonstrates a strong negative correlation between effective area (< 10 cm²) and current density. Once the effective area surpasses 10 cm², any changes no longer affect the current density significantly. Additionally, there exists a positive correlation between voltage and current density ; an increase in voltage leads to an increase in current density. Specifically, with a voltage greater than 1.5 V, cathode plate thickness less than 10 μm, and effective area less than 4 cm²; electrolysis efficiency can reach higher levels facilitating H₂ production.

According to Fig. 15a and b, a three-dimensional partial dependence relationship between voltage, temperature, and effective area on the rate of H₂ production is demonstrated. It can be observed that electrolysis temperature and voltage have the most significant influence on the rate of H₂ production, exhibiting a strong positive correlation. Fig. 15a reveals that as voltage increases from 0.9 V to 1.6 V and temperature rises from 650 °C to 850 °C, the rate of H₂ production escalates from 0.1 L/(h·cm²) to 0.55 L/(h·cm²). This indicates that increasing voltage and temperature could substantially improve the rate of H₂ production by

promoting internal chemical reactions in SOECs and accelerating electron transfer during electrolysis. As shown in Fig. 15b, there is a negative correlation between hydrogen production rate and effective area. For effective areas below 4 cm², an increase in the effective area leads to a decrease in the H₂ production rate. However, beyond 4 cm², no further changes are observed in the H₂ production rate. At a temperature of 850 °C, an increase in effective area results in a rapid decline of H₂ production rate from 0.47 L/(h·cm²) to below 0.35 L/(h·cm²). Therefore, the optimal solution is to achieve a higher level of electrolysis efficiency and easier production of hydrogen gas when the temperature is greater than 850 °C, the voltage is greater than 1.5 V, and the effective area is less than 4 cm².

3.6. Multi-objective optimization of SOEC process based on GA

To achieve a maximum H₂ production rate with the minimum current density of SOEC, the most important parameters determined by SHAP analysis are determined using the genetic algorithm combined with the TOPSIS method. By collecting the best solution points of the system to form a Pareto boundary curve, a set of optimization results is obtained for each optimized data point located at the forefront of Pareto analysis as shown in Fig. 16. The Pareto boundary curve demonstrates a positive correlation between the increase in current density and the significant enhancement of H₂ production rate. Finally, the TOPSIS method is

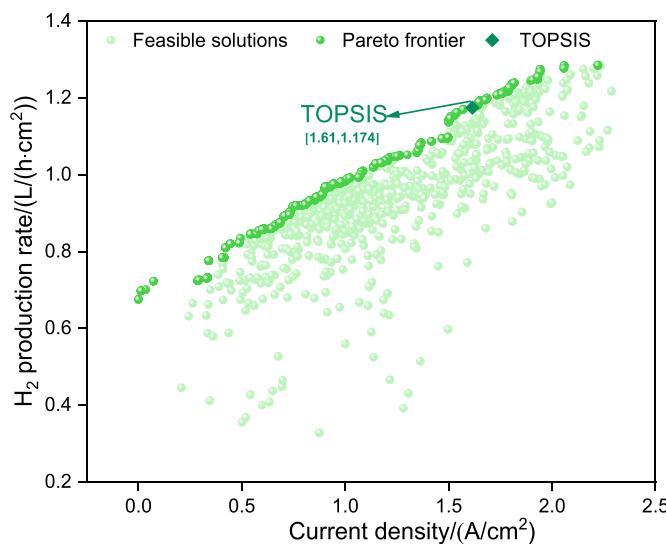


Fig. 16. Pareto curve of the multi-objective optimization of the SOEC process.

Table 5
Optimal value of the main parameters of the SOEC process.

Features	Optimal value
OE	LSM (Lanthanum Strontium Manganite)
Area (cm ²)	0.35
Electrolyte	YSZ (Yttria Stabilized Zirconia)
ET (μm)	10.01
Humidity (%)	34.61
Voltage (V)	1.56
Flow (mL/min)	46.8
T (°C)	843.95
GC	CO ₂ + H ₂ O + H ₂
P (kPa)	2.38
CO ₂	0.1
H ₂	0.2
H ₂ O	0.7

employed to select the optimal value, which indicates that the optimal current intensity and H₂ production rate are 1.61 A/cm² and 1.174 L/(h·cm²) with the optimal parameters listed in Table 5.

4. Conclusion

This study employs four ML models (including RF, DNN, SVR, and XGBoost) to predict the Ohmic resistance, current density, and H₂ production rate of the SOEC process by incorporating electrolytic cell descriptors, operating parameters, and feed conditions. The optimal model is explained through SHAP and PDP analyses and applied to optimize the key parameters of SOEC coupled with the genetic algorithm. The results demonstrate that the established dataset and selected input features are suitable for constructing ML models for SOEC since the correlation coefficient between the majority of input features is less than 0.5. The optimized XGBoost model is the most suitable for the SOEC process due to its highest R² value (R² = 0.95–0.97) and lowest RMSE across all three objectives compared with the RF, SVR, and DNN models. Furthermore, the SHAP method is performed to explain and analyze the predictive performance of the improved XGBoost model, revealing that the electrolytic cell descriptors are the most important for the model's prediction, contributing to 54.5%; followed by the operating parameters and feed conditions, 37.3%, and 8.2%. In particular, the effective area, voltage, and temperature of the SOEC have the greatest impact on the three output results, accounting for 21.6%, 16.6%, and 13.0%, respectively. Moreover, the interaction relationship between the most important input parameters of various output indicators is investigated by conducting a PDP analysis. It was found that at a temperature of around 850 °C, a

voltage of around 1.5 V, and an effective area of less than 4 cm², is relatively beneficial to produce hydrogen. Finally, the XGBoost model coupled with the GA algorithm is proposed to achieve the maximum H₂ production rate with the lowest current density in the SOEC process, which indicates the optimal current density and the H₂ production rate are 1.61 A/cm² and 1.174 L/(h·cm²). However, it should be validated by subsequent experiments.

CRediT authorship contribution statement

Qingchun Yang: Writing – review & editing, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Lei Zhao:** Writing – original draft, Resources, Methodology, Investigation, Data curation. **Jingxuan Xiao:** Software, Resources, Investigation, Data curation. **Rongdong Wen:** Visualization, Validation, Resources. **Fu Zhang:** Software, Resources, Data curation. **Dawei Zhang:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interests

The author declares the following financial interests/personal relationships which may be considered as potential competing interests: Jingxuan Xiao is currently employed by East China Engineering Science and Technology Co. Ltd. The other 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.

Acknowledgements

The authors are grateful for financial support from the National Natural Science Foundation of China (No. 22108052) and the High-end chemicals and cutting-edge new materials Technology Innovation Center of Hefei (HCHC202309).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gce.2024.04.004>.

References

- [1] A. Odenweller, F. Ueckerdt, G.F. Nemet, M. Jensterle, G. Luderer, Probabilistic feasibility space of scaling up green hydrogen supply, *Nat. Energy* 7 (2022) 854–865.
- [2] G. Squadrito, G. Maggio, A.A. Nicita, The green hydrogen revolution, *Renew. Energy* 216 (2023) 119041.
- [3] D. Guan, B. Wang, J. Zhang, R. Shi, K. Jiao, L. Li, Y. Wang, B. Xie, Q. Zhang, J. Yu, Y. Zhu, Z. Shao, M. Ni, Hydrogen society: from present to future, *Energy Environ. Sci.* 16 (2023) 4926–4943.
- [4] A. Hauch, R. Küngas, P. Blennow, A.B. Hansen, J.B. Hansen, B.V. Mathiesen, M.B. Mogensen, Recent advances in solid oxide cell technology for electrolysis, *Science* 370 (2020) eaba6118.
- [5] S. Zhang, N. Zhang, R. Smith, W. Wang, A zero carbon route to the supply of high-temperature heat through the integration of solid oxide electrolysis cells and H₂-O₂ combustion, *Renewable Sustainable Energy Rev.* 167 (2022) 112816.
- [6] D. Penchini, G. Cinti, G. Discepoli, U. Desideri, Theoretical study and performance evaluation of hydrogen production by 200 W solid oxide electrolyzer stack, *Int. J. Hydrogen Energy* 39 (2014) 9457–9466.
- [7] S. Biswas, G. Kaur, G. Paul, S. Giddey, A critical review on cathode materials for steam electrolysis in solid oxide electrolysis, *Int. J. Hydrogen Energy* 48 (2023) 12541–12570.
- [8] D. Jang, J. Kim, D. Kim, W.-B. Han, S. Kang, Techno-economic analysis and Monte Carlo simulation of green hydrogen production technology through various water electrolysis technologies, *Energy Convers. Manage.* 258 (2022) 115499.
- [9] IEA (2023), Global Hydrogen Review 2023, IEA, Paris, 2023. <https://www.iea.org/reports/global-hydrogen-review-2023>.
- [10] B. Zhang, N.F. Harun, N. Zhou, J.J. Colon-Rodriguez, D. Oryshchyn, L. Shadle, D. Tucker, S. Bayham, A real-time multiphysics model of a pressurized solid oxide electrolysis cell (SOEC) for cyber-physical simulation, *Energy Convers. Manage.* 298 (2023) 117778.
- [11] S. Song, X. Xiong, X. Wu, Z. Xue, Modeling the SOFC by BP neural network algorithm, *Int. J. Hydrogen Energy* 46 (2021) 20065–20077.

- [12] M. Ni, 2D thermal modeling of a solid oxide electrolyzer cell (SOEC) for syngas production by H₂O/CO₂ co-electrolysis, *Int. J. Hydrogen Energy* 37 (2012) 6389–6399.
- [13] V. Menon, V.M. Janardhanan, O.A. Deutschmann, Mathematical model to analyze solid oxide electrolyzer cells (SOECs) for hydrogen production, *Chem. Eng. Sci.* 110 (2014) 83–93.
- [14] S. Ali, K. Sørensen, M.P. Nielsen, Modeling a novel combined solid oxide electrolysis cell (SOEC) - biomass gasification renewable methanol production system, *Renew. Energy* 154 (2020) 1025–1034.
- [15] R. Yin, L. Sun, A. Khosravi, M. Malekan, Y. Shi, Control-oriented dynamic modeling and thermodynamic analysis of solid oxide electrolysis system, *Energy Convers. Manage.* 271 (2022) 116331.
- [16] C. Zhang, Q. Liu, Q. Wu, Y. Zheng, J. Zhou, Z. Tu, S.H. Chan, Modelling of solid oxide electrolyser cell using extreme learning machine, *Electrochim. Acta* 251 (2017) 137–144.
- [17] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: recent applications and prospects, *npj Comput. Mater.* 3 (2017) 54.
- [18] A. Mohamed, H. Ibrahim, R. Yang, K. Kim, Optimization of proton exchange membrane electrolyzer cell design using machine learning, *Energies* 15 (2022) 6657.
- [19] M. Raeesi, S. Changizian, P. Ahmadi, A. Khoshnevisan, Performance analysis of a degraded PEM fuel cell stack for hydrogen passenger vehicles based on machine learning algorithms in real driving conditions, *Energy Convers. Manage.* 248 (2021) 114793.
- [20] M.M. Kabir, S.K. Roy, F. Alam, S.Y. Nam, K.S. Im, L. Tijing, H.K. Shon, Machine learning-based prediction and optimization of green hydrogen production technologies from water industries for a circular economy, *Desalination* 567 (2023) 116992.
- [21] G. Bilgiç, E. Bendes, B. Öztürk, S. Atasever, Recent advances in artificial neural network research for modeling hydrogen production processes, *Int. J. Hydrogen Energy* 48 (2023) 18947–18977.
- [22] J. Li, L. Li, Y.W. Tong, X. Wang, Understanding and optimizing the gasification of biomass waste with machine learning, *Green Chem. Eng.* 4 (2023) 123–133.
- [23] S. Zhang, Y. He, Z. Zhang, C. Zhong, S. Zhang, Y. He, Z. Zhang, C. Zhong, Machine learning aided investigation on the structure-performance correlation of MOF for membrane-based He/H₂ separation, *Green Chem. Eng.* (2024), <https://doi.org/10.1016/j.gce.2024.01.005>.
- [24] N. Wang, Y. Guo, L. Liu, S. Shao, Numerical assessment and optimization of photovoltaic-based hydrogen-oxygen Co-production energy system: a machine learning and multi-objective strategy, *Renewable Energy* 227 (2024) 120483.
- [25] D. Vila, E. Hornberger, C. Toigo, Machine learning based state-of-charge prediction of electrochemical green hydrogen production: Zink-Zwischenschritt-Elektrolyseur (ZZE), *Energy and AI* 16 (2024) 10035.
- [26] M. Suvarna, T.P. Araújo, J. Pérez-Ramírez, A generalized machine learning framework to predict the space-time yield of methanol from thermocatalytic CO₂ hydrogenation, *Appl. Catal. B Environ.* 315 (2022) 121530.
- [27] M. Ahsan, M. Mahmud, P. Saha, K. Gupta, Z. Siddique, Effect of data scaling methods on machine learning algorithms and model performance, *Technologies* 9 (2021) 52.
- [28] P. Liashchynskyi, P. Liashchynskyi, Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS, 2019. <https://doi.org/10.48550/arXiv.1912.06059>.
- [29] H. Mai, T.C. Le, D. Chen, D.A. Winkler, R.A. Caruso, Machine learning for electrocatalyst and photocatalyst design and discovery, *Chem. Rev.* 122 (2022) 13478–13515.
- [30] S. Ascher, X. Wang, I. Watson, W. Sloan, S. You, Interpretable machine learning to model biomass and waste gasification, *Bioresour. Technol.* 364 (2022) 128062.
- [31] J. Li, L. Pan, M. Suvarna, X. Wang, Machine learning aided supercritical water gasification for H₂-rich syngas production with process optimization and catalyst screening, *Chem Eng. J.* 426 (2021) 131285.
- [32] Q. Yang, Y. Fan, J. Zhou, L. Zhao, Y. Dong, J. Yu, D. Zhang, Machine learning-aided catalyst screening and multi-objective optimization for the indirect CO₂ hydrogenation to methanol and ethylene glycol process, *Green Chem.* 25 (2023) 7216–7233.
- [33] X. Wan, Z. Zhang, W. Yu, Y. Guo, A density-functional-theory-based and machine-learning-accelerated hybrid method for intricate system catalysis, *Mater. Rep.: Energy* 1 (2021) 100046.
- [34] Y. Liu, W. Yan, S. Han, H. Zhu, Y. Tu, L. Guan, X. Tan, How machine learning predicts and explains the performance of perovskite solar cells, *Sol. RRL* 6 (2022) 2101100.
- [35] M. Kusnezoff, N. Trofimenko, M. Müller, A. Michaelis, Influence of electrode design and contacting layers on performance of electrolyte supported SOFC/SOEC single cells, *Materials* 9 (2016) 906.