論文 / 著書情報
Article / Book Information

Title	Machine learning-driven prediction and optimization of selective glycerol electrocatalytic reduction into propanediols
Authors	Muhammad Harussani Moklis, Cries Avian, Cheng Shuo, Sasipa Boonyubol, Jeffrey S. Cross
Citation	Journal of Electroanalytical Chemistry, Vol. 988, , pp. 119150
Pub. date	2025, 7
DOI	https://dx.doi.org/10.1016/j.jelechem.2025.119150
Creative Commons	Information is in the article.

Journal of Electroanalytical Chemistry 988 (2025) 119150

Contents lists available at ScienceDirect

Journal of Electroanalytical Chemistry

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

Machine learning-driven prediction and optimization of selective glycerol electrocatalytic reduction into propanediols

Muhammad Harussani Moklis [a,*], Cries Avian [b,c], Cheng Shuo [a], Sasipa Boonyubol [a], Jeffrey S. Cross [a]

[a] Energy Science and Engineering, Department of Transdisciplinary Science and Engineering, Institute of Science Tokyo, 2-12-1, Ookayama, Meguro-ku, Tokyo 152-8550, Japan
[b] Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan
[c] Department of Electrical Engineering, Universitas Brawijaya, Malang, Jawa Timur 65145, Indonesia

ARTICLE INFO

Keywords:
Glycerol upgrading
Electroreduction
Machine learning
XGBoost-PSO
Electrochemical conversion
Propanediols

ABSTRACT

Electrochemical conversion of crude glycerol–a surplus by-product of biodiesel production–into value-added propanediols (PDO) presents a sustainable bioresource valorization. However, optimizing selective glycerol electrocatalytic reduction (ECR) remains challenging due to complex interactions among multiple reaction parameters. Here, we employ an integrated machine learning-driven optimization framework combining XGBoost with particle swarm optimization (PSO) to predict and optimize glycerol ECR performance, targeting both conversion rate (CR) and electroreduction product yields (ECR PY). A dataset of 446 experimental datapoints curated from published literature was used to train the XGBoost model, achieving high prediction accuracy (R^2 of 0.98 for CR; 0.80 for ECR PY), outperforming other algorithms and demonstrating robustness against unbalanced datasets. Feature analysis revealed that low-pH electrolytes and longer reaction times significantly enhance both outputs, while higher temperatures and carbon-based electrocatalysts positively influence ECR PY by facilitating C—O bond cleavage in glycerol. XGBoost-PSO optimization predicted maximum CR (100 %) using a Pt cathode at 24.15 h, 24.66 °C, pH 1.08, 66.96 rpm stir rate, 0.43 M electrolyte concentration, and 0.28 A/cm^2 current density. Meanwhile, the highest ECR PY (53.29 %) was predicted with a carbon cathode at 22.27 h, 78.87 °C, pH 0.99, 650.18 rpm, 3.84 M electrolyte, and 0.14 A/cm^2. Experimental validation confirmed the model's predictive accuracy within ~10 % error. GC–MS further validated the selective formation of PDOs, with yield of 21.01 % under optimized conditions. This framework offers a robust, data-driven alternative to traditional trial-and-error approaches, providing mechanistic insights and practical guidance for scalable, economically viable glycerol ECR in biodiesel industry.

1. Introduction

According to the International Energy Agency (IEA), global biodiesel generation, as of 2021, is recorded exceeding 50 billion liters, generating approximately 5 billion liters of crude glycerol [1,2], typically accounting for 10–20 % of the total volume. This abundance of crude glycerol [3] has spurred increased research interest in effectively converting this waste product into higher-value biochemicals or biofuels [4], thus, can significantly improve the overall profitability and sustainability of the biodiesel industries. One promising approach is electrocatalytic reduction (ECR), a selective electrochemical conversion reaction that can efficiently convert glycerol into various value-added biochemicals [5,6] such as propanediols (PDOs), propanols (POHs), ethylene glycol, and other valuable compounds widely utilized in biofuels, solvents, pharmaceutical and polymer industries.

ECR process is favored for its high efficiency, ease of manufacturing setup, environmental friendliness, and simplicity in its system integration [7]. Based on previous studies, this electrochemical approach promotes effective deoxygenation reaction of removing oxygen content from oxygenated compounds [8,9]. Despite its potential, selective glycerol ECR remains underdeveloped due to challenges in optimizing reaction conditions and electrocatalyst performance. Current studies [10–17] have primarily focused on experiments using pure aqueous glycerol under varying reaction conditions. Therefore, achieving an efficient selective ECR of glycerol to PDOs, both in terms of performance

* Corresponding author.
E-mail address: harussani.m.ab@m.titech.ac.jp (M.H. Moklis).

https://doi.org/10.1016/j.jelechem.2025.119150
Received 8 February 2025; Received in revised form 21 April 2025; Accepted 22 April 2025
Available online 25 April 2025

Nomenclature

PDO	Propanediol
POH	propanol
ECR	electrocatalytic reduction
ML	machine learning
CR	glycerol conversion rate
ECR PY	electroreduction product yields
R^2	R-squared score
PSO	particle swarm optimization
DT	decision tree
RF	random forest
SVM	support vector machine
ANN	artificial neural network
GPR	Gaussian process regression
XGBoost	extreme gradient boost
kNN	k-nearest neighbour
IEA	International Energy Agency
LIME	Local Interpretable Model-agnostic Explanations
SHAP	SHapley Additive exPlanations
PI	permutation importance
GC–MS	gas chromatography mass spectrometry
XAI	Explainable artificial intelligence
Mi	initial concentration of glycerol
T	processing temperature
J	current density
V	applied potential
I	electric current
t	pH value of the reaction medium
pH	operating time
M	molarity
rpm	rotation per minute
A/cm^2	ampere per square centimeter
$^{\circ}C$	degree Celsius
hrs	hours
mg/mL	milligrams per milliliter
μm	micrometer
mm	millimeter
mL	milliliter
V	volt
cm^2	square centimeter
y_i	real experimental result
y_m	expected result of the output
$\bar{y}$	average of the output values
n	number of samples
R^2_{train}	R^2 value for training set
R^2_{test}	R^2 value for test set
$RMSE_{test}$	root-mean-square error for test set
MAE_{test}	mean absolute error for test set
C–O	carbon-oxygen single bond
CBAC	carbon black/activated carbon
CBD	carbon black diamond
Pb	lead
Pt	platinum
SS	stainless steel
$TiRuO_2$	titanium ruthenium oxide
Zn	zinc
H_2SO_4	sulfuric acid
HCl	hydrochloric acid
KCl	potassium chloride salt
KCl/HCl	acidified KCl
Na_2SO_4	sodium sulfate salt
NaCl	sodium chloride salt
NaCl/HCl	acidified NaCl salt
NaOH	sodium hydroxide
CV	cyclic voltammetry
CP	chronopotentiometry
CA	chronoamperometry
GAN	generative adversarial networks

and economy, hinges on optimizing the selective ECR reaction and designing an electrocatalytic reactor with high-selectivity electrocatalyst fabrication. This is considered the primary obstacle and opportunity in the near future for the implementation of ECR technology in the treatment of real crude glycerol and even for other oxygenated waste compounds.

Current investigations typically use pure aqueous glycerol reagent under ambient processing temperatures with various combinations of other operating parameters–in electrolytes with different pH values (acidic, alkaline and near neutral media) [18–20], applied voltage and current density [10,11], residence time [15] as well as different types and materials of working electrodes (cathodes) [10–12,21], which act as electrocatalysts in the reaction. However, the majority of the reported data shows inconsistencies which due to diverse experimental setups, unregulated operational conditions as and the ambiguous nature of the electrocatalysts upon operating under these traditionally determined operational conditions. Hence, this inconsistency complicates the understanding of ECR mechanisms and selectivity. To resolve this significant knowledge gap, systematic data-driven research on processing parameter optimization is paramount to advancing the field, particularly for real crude glycerol applications.

Traditionally, research related to parameter optimization for system design–including glycerol ECR [6,14,22]–has relied on trial-and-error experiments across laboratory, pilot, and industrial scales–an approach that is costly, labor-intensive, and time-consuming, ultimately hindering technological advancement [23]. Machine learning (ML), specifically through the concept of target-oriented inverse design, offers a systematic, data-driven approach to predict reaction outcomes and optimize process parameters for improved electrochemical efficiency and product yields [24–26]. ML has also been widely applied in other fields, including engine performance prediction [27,28], electrolysis modeling [29–31], and pollution forecasting [32]. Drawing upon decades of studies that has generated substantial experimental data, utilizing ML becomes a viable toolkit for making predictions based on experiential knowledge by uncovering complex patterns, identifying nonlinear interactions, and accurately predicting output using a comprehensive array of input features from existing literature data. Despite its wide-ranging success in other domains, the application for optimization in glycerol ECR remains largely underexplored, presenting a promising opportunity for accelerating progress in this field.

Here we aim to predict and optimize the electrochemical performance–specifically, the conversion rate (CR) and electroreduction product yield (ECR PY) of PDOs–for selective glycerol ECR reaction by developing an ML-guided optimization framework. By employing a dataset of 446 experimental data points from previous studies, we built predictive models for CR and ECR PY that uncover key processing parameters. These models were integrated with particle swarm optimization (PSO) to systematically determine optimal reaction conditions, which were then experimentally validated to confirm the model's accuracy. This data-driven approach demonstrates how integrating ML-driven inverse design can revolutionize bioresource utilization by replacing traditional trial-and-error methods, resolving data inconsistencies, reducing dependency on resource-intensive experimentation, and accelerating process development. To the best of our knowledge, this is the first study

to apply ML-guided inverse design for systematic and accelerated development of glycerol ECR optimization, bridging the gap between experimental studies and scalable industrial applications while addressing data inconsistencies in prior studies. Our findings provide a foundation for future advancements in electrochemical bioresource utilization and sustainable glycerol valorization.

2. Research methodology

We developed an ML-guided framework to predict and optimize the selective ECR of glycerol to propanediols, integrating data-driven modeling with bio-inspired optimization framework with experimental validation. As illustrated in Fig. 1, our approach mainly involves four key steps: (1) extracting and preprocessing 446 experimental data points from literature, (2) developing multiple ML models to predict CR and ECR PY, (3) optimizing reaction conditions using extreme gradient boost (XGBoost) model integrated with PSO for inverse design, and (4) validating the optimized reaction conditions via electrochemical experiments. Further methodological details are provided in the following subsections.

2.1. Extraction and preprocessing of data sets

2.1.1. Data extraction and standardization

The electrochemical performance data (CR and ECR PY) as well as its corresponding operating parameters of the selective ECR of glycerol to PDOs were extracted from the published literature [10–12,15,33,34]. Various main processing parameters–initial glycerol concentration, cell type, stirring rate, type of anode and cathode electrodes, their surface area, processing temperature, applied voltage and current density, current, reaction time, type of electrolytes and its pH value–were collected and compiled into a single CSV file. Since electrochemical efficiency was highly parameter-dependent [5,35], these parameters were implemented as input features within the predictive model, while total glycerol CR, and ECR PY (1,2- and 1,3-PDO yields) served as output labels. The compiled dataset was standardized into consistent formats and SI units, addressing the variations and data inhomogeneity across different studies–various kind of experimental set up, the way the equipment being handled, how the sample being measured, etc. A summary of extracted features and labels is provided in Table S1, with data distribution illustrated in Fig. S1.

2.1.2. Missing value imputation and categorical feature encoding

The inconsistencies of the literature can also be explained by the presence of missing values. These values, rather than being discarded, were imputed using k-nearest neighbours (kNN) technique (k = 10 employed in this study), which proved the most suitable with our datasets. Similarly, encoding step is needed to retain the categorical information, provided by the categorical features within the dataset, while enhancing models training. These features–the electrode type, electrolyte type, and the different cell configurations (double or single cells)–were dummy-encoded into numerical values of a set of binary variables (0 or 1 for absence or presence, respectively). Dummy encoding uses N binary variables for categories in a variable [36].

2.1.3. Input feature selection and data normalization

Spearman's rank correlation analysis was applied to provide a preliminary understanding of the degree of relationship between features, with highly correlated pairs (>0.90 values) filtered to remove redundancy. Subsequently, permutation importance (PI) was also carried out to refine feature selection. Finally, data normalization using a standard scaler was utilized to ensure consistent value ranges and improve model performance.

Following these preprocessing stages, the current dataset contained 446 data points, and 32 features (31 input features including categorical and numerical features, only 1 output feature–CR or ECR PY). To avoid data leaking, the training and test sets were kept entirely separate.

2.2. Developing and evaluating of ML modeling

2.2.1. ML model selection – hyperparameter tuning and training

To investigate and discover the best prediction model for glycerol CR and ECR PY, six main ML algorithms–decision tree (DT), random forest (RF), XGBoost, support vector machine (SVM), Gaussian process regression (GPR) and artificial neural network (ANN)–which usually employed in chemistry-related research were trained and tested. These models were coded and programmed using scikit-learn in Google Colab (Python, macOS environment). The compiled dataset was split into training and test sets with ratio of 7:3 for model development. To achieve the best architecture for each model, grid search technique was employed for algorithm's hyperparameter tuning. A detailed description regarding the hyperparameter tuning for each model is given in Table S2.

2.2.2. Model evaluation and performance metrics

To avoid overfitting and underfitting issues [37], 10-fold cross-validation method was applied. Model performance was measured by the average score from the 10 iterations of the evaluation metrics–R-squared (R^2) score, root-mean-square error (RMSE) and mean absolute error (MAE). Here, overfitting can be detected by identifying a significant decline in performance on the test folds compared to the training folds, which were computed as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_m)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

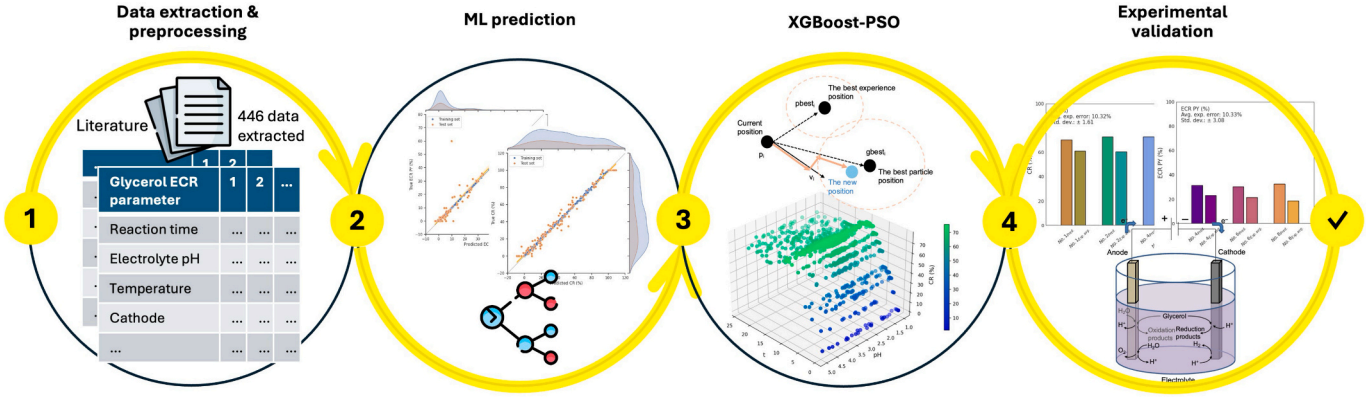

Fig. 1. Workflow of the ML-guided prediction and optimization for selective ECR of glycerol to propanediols (PDOs).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_m)^2}{n}}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \bar{y}|$$

where y_i, y_m and $\bar{y}$ denotes the real experimental result, expected result of the output, and the average of the output values, respectively. Whereas n represents the number of training, test, and cross-validation test samples. R^2 score indicates how well the model explains data variation and measures how well the predicted values match the actual data. RMSE emphasizes larger errors by averaging squared differences between predicted and actual values, while MAE provides a straightforward average of the absolute differences between predictions and actual values.

In essence, higher R^2 scores and lower error values indicate better model accuracy. The best-performed algorithm was then employed as our main ML model for further works of output prediction and inverse design.

2.2.3. ML-assisted feature exploration–LIME and SHAP approaches

Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) approaches were integrated with ML models to visualize the effect of various input features on the selective glycerol ECR and how do they arrive to their predictions. This method was utilized to capacitate the lack of interpretability for most of ML "black-box" models [38,39].

2.3. Implementing inverse design of XGBoost-PSO

Implementation of inverse design in ML is paramount to visualize the knowledge from the well-trained model, which can then be applied to pragmatic experimental tests. In this study, PSO algorithm, embedded with the XGBoost model, was used to search for best combination of optimal input features for maximum CR and ECR PY. The XGBoost-PSO workflow is summarized in Fig. S2. To enable experimental validation of the yielded results, only easily manipulable reaction parameters were taken into account.

2.4. Experimental validation

2.4.1. Electrochemical testing

In order to verify ML predictions, potentiostatic-mode electrolysis was conducted using a Daiwa SS-330 W potentiostat (Daiwa, Japan) in a 20-mL single-cell reactor. A Pt and carbon rod served as the counter and working electrode, respectively, in a two-electrode setup, see Fig. 2.

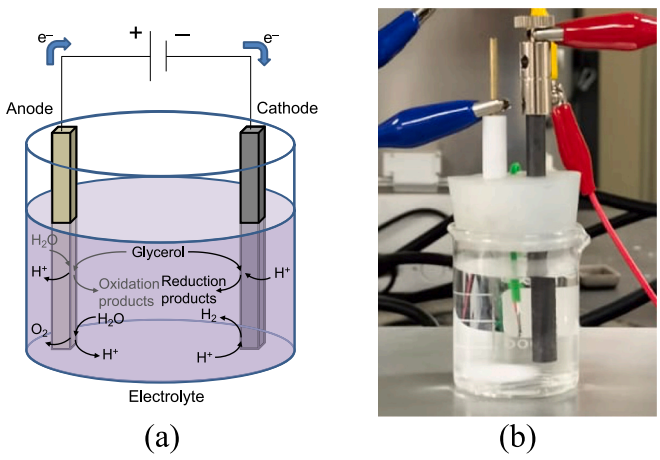

Fig. 2. (a) Proposed schematic of selective ECR of glycerol, and (b) experimental setup used.

2.4.2. Liquid product characterization–GC–MS analysis

Target electroreduction products (1,2- and 1,3-PDOs) and glycerol CR were analyzed using a gas chromatography mass spectrometry (GC–MS) (QP2010, Shimadzu, Japan) with SH-200MS capillary column (30 m × 0.25 mm × 0.25 μm). Samples underwent neutralization to pH 7 with sodium hydroxide, filtered (0.45 μm nylon syringe), and dissolved in 10 mL of methanol, acts as solvent for dissolution, per 0.01 g of the filtered sample. The injection volume for GC–MS was set at 2 μL. The GC–MS temperature program was set as follows–the oven temperature initially began at 50 °C for a duration of 5 min before gradually rising to 250 °C at a heating rate of 5 °C/min, followed by a final hold time of 10 min.

Subsequently, the detected peak of compounds was compared with those in the NIST MS library. Integrated peak areas from GC–MS analysis were converted into compound concentrations using standard calibration curves for glycerol and PDOs, which were established with known reference concentrations. The initial and final concentrations were determined based on these calibration curve equations. Glycerol CR and ECR PY were calculated using these equations.

$$\mathrm{CR}\ (\%) = \frac{\text{Concentration of converted glycerol (mg/mL)}}{\text{Initial concentration of glycerol (mg/mL)}} \times 100$$

$$\mathrm{ECR\ PY}\ (\%) = \frac{\text{Concentration of ECR product (PDO) (mg/mL)}}{\text{Initial concentration of glycerol (mg/mL)}} \times 100$$

Here, the converted glycerol concentration was determined by the differences between initial and the final glycerol concentration after the reaction.

3. Results and discussion

3.1. Descriptive statistical analysis of the data set

Descriptive analysis in ML is necessarily helping to summarize and understand the main characteristics of the dataset. This analysis includes a variety of tasks such as basic summary statistics (mean, median, mode, minimum and maximum for numerical features), correlation analysis, and feature selection to improve predictive accuracy.

Fig. 3 displays the Spearman rank correlation matrix, highlighting relationships between reaction parameters. We could observe that there was strong positive correlation between current flow of *I* and *Cathode surface area*, demonstrated by higher correlation value of 0.89. Similarly, *Electrolyte_H2SO4* and *Feed_Enriched crude glycerol* pair as well as *Cell type_Double cell* and *T* pair exhibited strong positive correlation with value of 0.88 and 0.87, respectively. To mitigate collinearity and improve model performance, PI analysis was carried out (Figs. S3(a) and (b)). The results revealed that compared to *I* feature, *V* exhibited lower PI scores for both labels, leading to its exclusion from further modeling to reduce feature redundancy. On the other hand, the other feature pairs demonstrated weaker positive correlation with Spearman rank coefficient lower than 0.80 as well as higher PI values. It shows that the remaining features exhibited relatively individual contribution to the prediction of the models.

Upon the completion of preprocessing the raw data and prior to modeling phase, the dataset was prepared for further descriptive analysis on both the input features and labels to glean its preliminary insights. A total of 32 input parameters were selected based on their direct influence on the electrochemical reaction, as illustrated in Fig. 3, including numerical features of glycerol initial concentration, stirring rate, electrodes' surface area, temperature, reaction time, applied potential and pH of electrolyte as well as the categorical features. Categorical features such as feed and cell types, anode and cathode, as well as different electrolytes were categorized accordingly for a higher degree of visualization due to the excessive number of elements (7 varieties of cathode and 8 kinds of electrolytes). These features were chosen due to their well-established roles in governing the reaction kinetics, mass

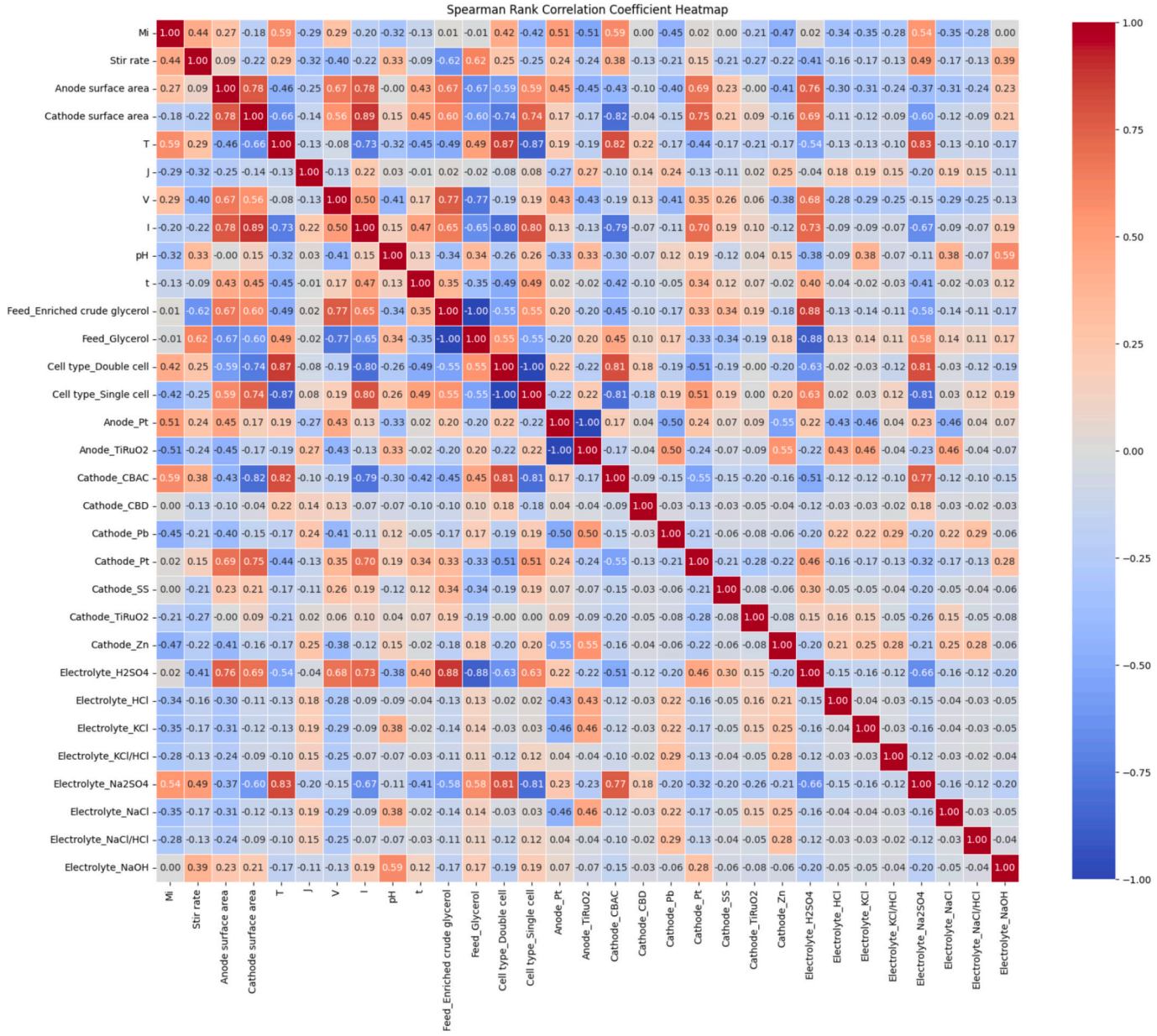

Fig. 3. Input feature correlation matrices from Spearman's rank coefficient analysis. The analysis used to measure monotonic correlation between two features, with ranges of −1 to 1, where input features with high pairwise correlation (>0.90) should be filtered out [40]. It may provide difference-blind information as they carrying essentially the same information to the modeling and could even introduce multicollinearity issues [41].

transport and selectivity of glycerol ECR.

Basic summary statistics of the cleaned preprocessed data set was carried out and the results displayed in Fig. S1. The distribution of output variables–CR and ECR PY–was displayed as according to its minimum (0.0 and 0.0), maximum (100.0 and 63.6), median (40.35 and 1.58), mean (44.54 and 6.25) and mode (100.0 and 0.0) values. These outputs were selected as key performance indicators because they quantitatively reflect both the extent of glycerol conversion and the efficiency of ECR towards desired products of PDOs. In addition to previous correlation analysis between features and labels, it is clear that the relationship between each of the inputs and outputs is evidently non-linear. This justified the use of ML algorithms capable of handling both linear and non-linear relationships [42], which made them more suitable for predicting the value of CR and ECR PY of the selective glycerol ECR reaction in this specific study.

3.2. Evaluation of ML model performance for CR and ECR PY

The initial phase of ML modeling involves the identification of the most suitable algorithms for predicting outputs in terms of the operating conditions of selective glycerol ECR. A comparative analysis of regression-type supervised ML algorithms was done based on their prediction performance (see Table 1). Herein, this step plays a pivotal role in achieving high-accuracy model development for the generation of precise predictions. This precision is critical for supporting subsequent inverse design processes within the context of glycerol ECR reactions.

Our study delved into the performance metrics of six widely used ML algorithms, including DT, RF, SVM, ANN, GPR and XGBoost. As aforementioned above, a thorough comparative analysis was conducted, assessing their R^2 scores on both training and test sets, along with evaluating their RMSE as well as MAE values based on two different continuous outputs, namely the conversion rate (CR (%)) and

Table 1
Comparison for performances (R^2_{train}, R^2_{test}, $RMSE_{test}$ and MAE_{test}) of different optimized models.

Algorithms	CR (%)				ECR PY (%)			
	R^2_{train}	R^2_{test}	$RMSE_{test}$	MAE_{test}	R^2_{train}	R^2_{test}	$RMSE_{test}$	MAE_{test}
DT	0.999	0.913	0.302	0.160	0.681	0.609	0.664	0.244
RF	0.993	0.948	0.234	0.141	0.969	0.789	0.488	0.184
SVM	0.965	0.965	0.192	0.3704	0.749	0.532	0.726	0.318
ANN	0.992	0.972	0.172	0.111	0.923	0.713	0.568	0.164
GPR	0.995	0.987	0.118	0.077	0.999	0.726	0.555	0.199
XGBoost	0.999	0.977	0.170	0.127	0.999	0.804	0.481	0.168

electroreduction product yields (ECR PY (%)). High R^2 scores as well as low RMSE and MAE values on the test sets were expected for robust predictive models.

As summarized in Table 1, for CR prediction, GPR model recorded the best prediction accuracy with the highest test set R^2 value (R^2_{test}) of 0.987, and the lowest test set RMSE ($RMSE_{test}$) and MAE (MAE_{test}) of 0.118 and 0.077, respectively. High predictive accuracy with small error values also has been recorded by previous electrochemical works using GPR algorithm [43,44]. XGBoost and ANN models also exhibited comparable high R^2_{test} scores of 0.977 and 0.972 with low error values. This trend aligns with previous studies, such as Hitt et al. [45], where ANN and gradient boosting models in ML prediction for alkaline hydrogen electrooxidation (R^2_{test} of 0.95 and 0.90, respectively). Consequently, published studies [40,46,47] consistently report advanced ML algorithms such as GPR, ANN and XGBoost demonstrated notably superior predictive performance, achieving >0.90 R^2 as compared with DT, RF and SVM models. Thus, highlighting the ability of these models to effectively capture and simulate intricate non-linear relationships within the dataset with higher adaptability.

Notably, in the context of Explainable AI (XAI), it is imperative to acknowledge that the XGBoost model showcases better interpretability features assisted by the gray box method of SHAP, which allows for the identification of key factors influencing CR and ECR PY values, surpassing both ANN and GPR in this regard [48]. This interpretability is essential for guiding experimental conditions in ECR optimization, offering insights that are easily communicated and actionable for experimental follow-up. In addition, XGBoost model tend to be simpler, using decision trees as base learners, which inherently have a hierarchical structure of gradient boosting framework that can be visualized and understand more easily [49,50] than the complex interconnected layers of neurons in ANN [23] and the computational complexity in GPR [43]. Next, compared to some more advanced algorithms, like ANN and GPR, XGBoost offers an excellent balance between predictive power and computational efficiency [40,45–47]. Given the complexity of the dataset and the goal of producing reliable, interpretable results, XGBoost provided an optimal solution without requiring extensive computational resources [48]. Thus, these serve as a justification for the preference of XGBoost over ANN and GPR in this study. Fig. 4 illustrates the prediction performance of well-trained XGBoost model as the best algorithm for the datasets.

Next, we expanded our prediction models by integrating comparable input features related to processing parameters alongside the second label of ECR PY (%). The optimal hyperparameters for each model trained in this phase were summarized in Table S2. As shown in Table 1, XGBoost model performed the most accurate predictions with the highest R^2_{test} (0.804) and the lowest $RMSE_{test}$ (0.481), followed by the RF model with R^2_{test} of 0.789 and $RMSE_{test}$ of 0.488 and GPR model with lower R^2_{test} of 0.726, consistent with Zaffar et al. [51], which observed XGBoost and RF models' high R^2 accuracy compared to other models. Conclusively, the results obtained appeared to be lower than the prediction with CR label. This finding highlights how different target labels significantly influence the prediction accuracy. The reduced performance by all algorithms were recorded which might be due to the presence of a significant portion of the outputs with zero values due to

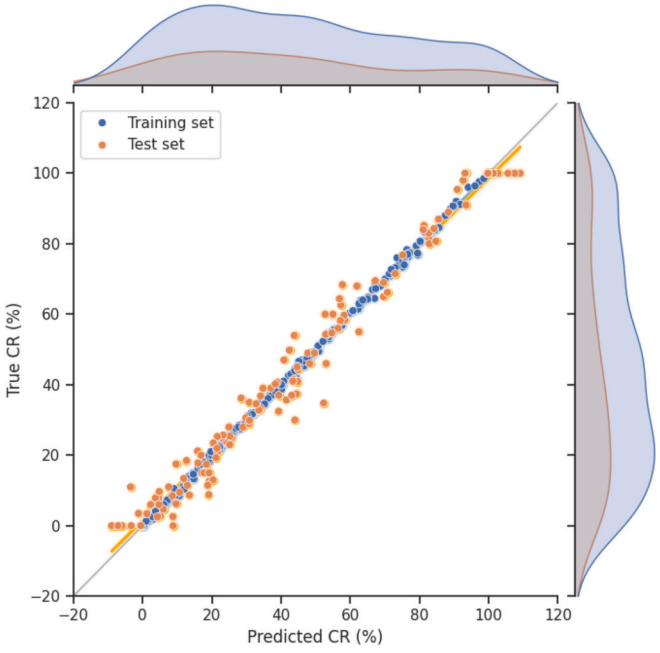

Fig. 4. Prediction performance of XGBoost model trained by glycerol electrocatalytic reduction reaction conditions as input features. The diagonal line represents the perfect prediction line (gray line). The distribution of output data in terms of true and predicted conversion rate values is exhibited in the upper and right side of each figure. Light blue color for the training set while orange color for the test set. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the zero ECR PY yield extracted from literature [52,53], resulting in a zero-inflated or skewed distribution, as shown in Fig. 5. Such class imbalance likely impaired the model's ability to capture the underlying patterns in the data, especially in algorithms sensitive to output distribution. Moreover, noisy or inconsistent labeling in the literature could have further negatively impact prediction accuracy [53–55]. These issues may cause models to be biased towards predicting zero values and thus, leading to poor generalization capability for non-zero outcomes.

However, we could observe that XGBoost is less prone to being affected by imbalanced datasets–particularly those containing many zero values in the output–compared to some other ML algorithms. The robustness of XGBoost algorithm with unbalanced dataset was also reported by previous literature with high predictive accuracy (R^2_{test} > 0.90) [40,56,57]. This might be due to that XGBoost is an ensemble learning algorithm that builds a series of decision trees sequentially, where the final prediction is a combination of predictions from multiple trees [58,59]. Such a structure reduces overfitting and helps mitigate the impact of class imbalance by capturing complex patterns across the dataset [55]. Moreover, XGBoost's gradient boosting mechanism allows it to assign greater weights to misclassified samples, improving its ability to generalize across diverse and sparse data spaces [52,55,59]. This indicates the reliability of the XGBoost model for predictive purposes,

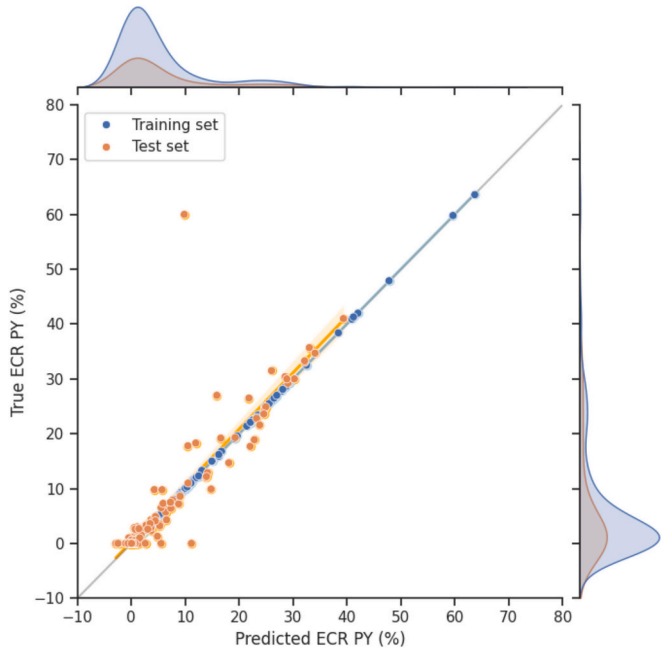

Fig. 5. Prediction performance of XGBoost model trained by glycerol ECR with its reaction condition as features and labels/outputs of electroreduction products–1,2- and 1,3-PDOs.

particularly in underdeveloped technology of glycerol ECR, where reported datasets are often limited in size, heterogenous, and skewed. Its consistent performance across both training and testing datasets further reinforces its reliability and predictive power. Hence, it provides stronger rationale for our continued utilization of the XGBoost model in subsequent prediction tasks.

3.3. Feature importance measure

To deepen the understanding of the outcome, the importance of input features was assessed. This methodology in feature engineering seeks to pinpoint the dominant processing parameters that significantly affect the selective glycerol ECR reaction. Similar to previous stages, feature selection utilizing PI was conducted. As explained by Gregorutti et la. [60], while PI generally used to measure the individual contribution of an input feature to the overall model prediction, it cannot clearly discriminate whether the feature show higher or lower importances towards the output. Conversely, there are reliable methods that assign local feature-importance scores for providing a mechanistic understanding of ML predictions, namely LIME and SHAP models [48]. Both analyses quantify individual contributions of input features as well as uncover whether the descriptors or features positively or negatively influence the target variables. Here, insights from both models used to better interpret the outcomes yielded from the XGBoost model.

3.3.1. Feature importance analysis for prediction of glycerol CR

Fig. 6 illustrates the significance of input features using mean absolute SHAP values as a measure of feature importance. Specifically, Figs. 6(a) and (b), the SHAP plots, display the global understanding of feature importance by visualizing the impact of each feature on model output. While Fig. 6(c), the waterfall plot, and Fig. 6(d) (using LIME method) provide local explanations for individual predictions. Synergistically, all plots present a holistic view of how the model come to the prediction.

Based on Figs. 6(a) and (b), *t* emerges as the most critical feature by positively correlated to CR output, with the mean absolute SHAP value of 18.0. Extended reaction durations were shown to enhance glycerol conversion rates, emphasizing the significant role of processing time as the foremost influential parameter in the efficiency of glycerol electrochemical conversion within the cell [10,61,62]. It was followed by *pH*, negatively correlated to the conversion rate of glycerol indicating that a low pH in the electrolyte (representing a strong acidic medium) enhances glycerol conversion. This observation aligns with findings from our previously published literature study [5], which discussed the positive impact of a higher concentration of H^+ ions in acidic electrolytes on electrolyte conductivity and conversion rates. *Cell type_Double cell* and *Cathode_Pt* are reported also as highly essential features with higher mean absolute SHAP values, and their absence (shown by 0 values in the waterfall plots) negatively influenced the prediction of CR. In line with the literature, the utilization of Pt as cathode positively correlated to higher feedstock conversion rate. This is due to their robustness and high catalytic properties. However, the drawback lies in its high cost, impeding its widespread utilization as electrocatalysts in the industrial application. Thus, presenting a significant knowledge gap in this research area [63]. Interestingly, *T* demonstrated a weaker influence on the conversion rate, suggesting that glycerol electrochemical conversion can occur effectively even under ambient temperature conditions.

Beyond individual parameter effects, combined interactions between multiple features played a role in CR optimization. For instance, despite near-neutral pH, glycerol achieved complete conversion (CR of 99.47 % and 102.20 %), primarily due to the favorable values of other parameters such as *Stir rate*, and *T*. Overall, these physical interpretations affirm the validity of our trained models as well as lay the groundwork resulted from ML-assisted analysis for our further study.

3.3.2. Feature importance analysis for prediction of ECR PY

In term of electroreduction product yields, as shown in Fig. 7, the most impactful features contributing most to ECR PY values within selective glycerol ECR with average SHAP value of 6.0 was *t*. Likewise, *T* also demonstrated strong positive correlation towards model output of electroreduction product yield, corresponding to its high negative impact value towards ECR PY at low *T* value as shown in Figs. 7(b) and (c), tailed by *pH* which negatively correlated to the product yields. Following that, input features such as *Stir rate* and *Cathode_CBAC* showed positive correlation with average impact value of 1.5 and 1.4, respectively. These insights were collectively agreed by different ML-assisted interpretability tools presented in Fig. 7(d).

Strong positive correlation recorded between *t* with ECR PY (Fig. 7 (b)) as well as glycerol CR (Fig. 6(b)) demonstrates that electrolysis of glycerol is a time-dependent reaction where longer processing time is needed to completely convert glycerol into various reduction products, as reported by numerous researchers. However, utilization of extended reaction time is a hindrance industrially. Thus, the study on the optimization of major influential features, such as association of operating temperature with electrocatalysts or cathode materials, towards the reaction and product yields is paramount to effectively reduce the time of reaction. *T* was positively correlated with ECR PY values, followed by the cathode surface area features which negatively correlated with the yield of reduction products. This significant finding demonstrates that higher operating temperatures lead to increased yield of electroreduction product–PDOs. This is due to that, at higher temperature, enhanced kinetic energy of reacting species within the electrolyte boosts collision frequency and interactions with catalytic sites, as noted by Kishore et al. [64], which lead to the breaking down of one C—O bond within glycerol. This reaction also can be called as electrodeoxygenation (Eq. 1). However, several studies also mentioned that excessive temperatures can lead to structural changes, deactivation, or degradation of the electrocatalyst [65–67].

$$C_3H_8O_3\,(glycerol) + 2H^+ + 2e^- \rightarrow C_3H_8O_2\,(Reduction\,products\text{–}PDOs) + H_2O \quad (1)$$

Hence, effective electrocatalyst activation occurs within specific temperature ranges, where molecule mobility on the electrode surface

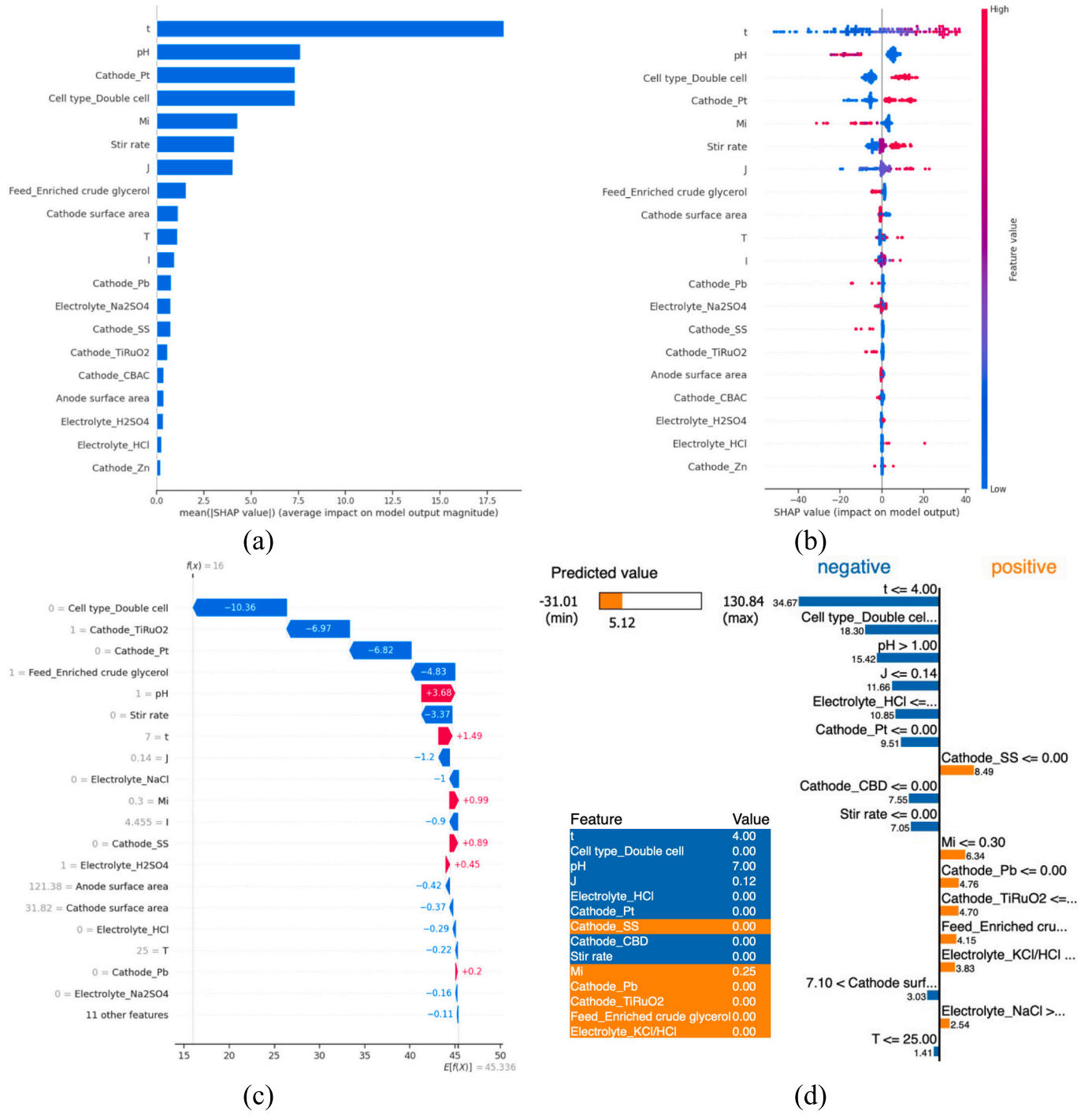

Fig. 6. ML-assisted statistical analysis for the XGBoost prediction model of glycerol ECR with CR labels. (a) Feature importance scores, and (b) SHAP value of input features' impact (summary plot) on the model output. Its effects were described by the difference of the color spectrum where high input feature values (in red) described as positive SHAP values, whereas low input feature values (in blue) labelled as negative SHAP values). (c) SHAP values as waterfall plots for the influence of the features towards the prediction of CR values. Note that the waterfall plots are designed to show explanations for individual predictions for the CR output at its specified value. (d) LIME methodology to determine the impact of each feature on the CR. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

increases, facilitating adsorption and desorption processes in electrocatalysis. The interplay of these factors underscores the importance of carefully controlling and optimizing the reaction temperature for electrocatalysts to achieve maximum catalytic reactivity, balancing kinetic and thermodynamic considerations for efficient and stable electrochemical reduction as well as electro-deoxygenation processes.

3.4. ML-based inverse design for glycerol ECR using XGBoost-PSO

To expedite advancements in glycerol ECR, implementing inverse design is vital. This crucial step serves to validate the predictions, addressing a significant knowledge gap in many published ML-based studies. This approach signifies a fundamental shift in perspective, commencing with the target reaction rate and then searching for the optimal operational parameters. Consequently, the utilization of a well-

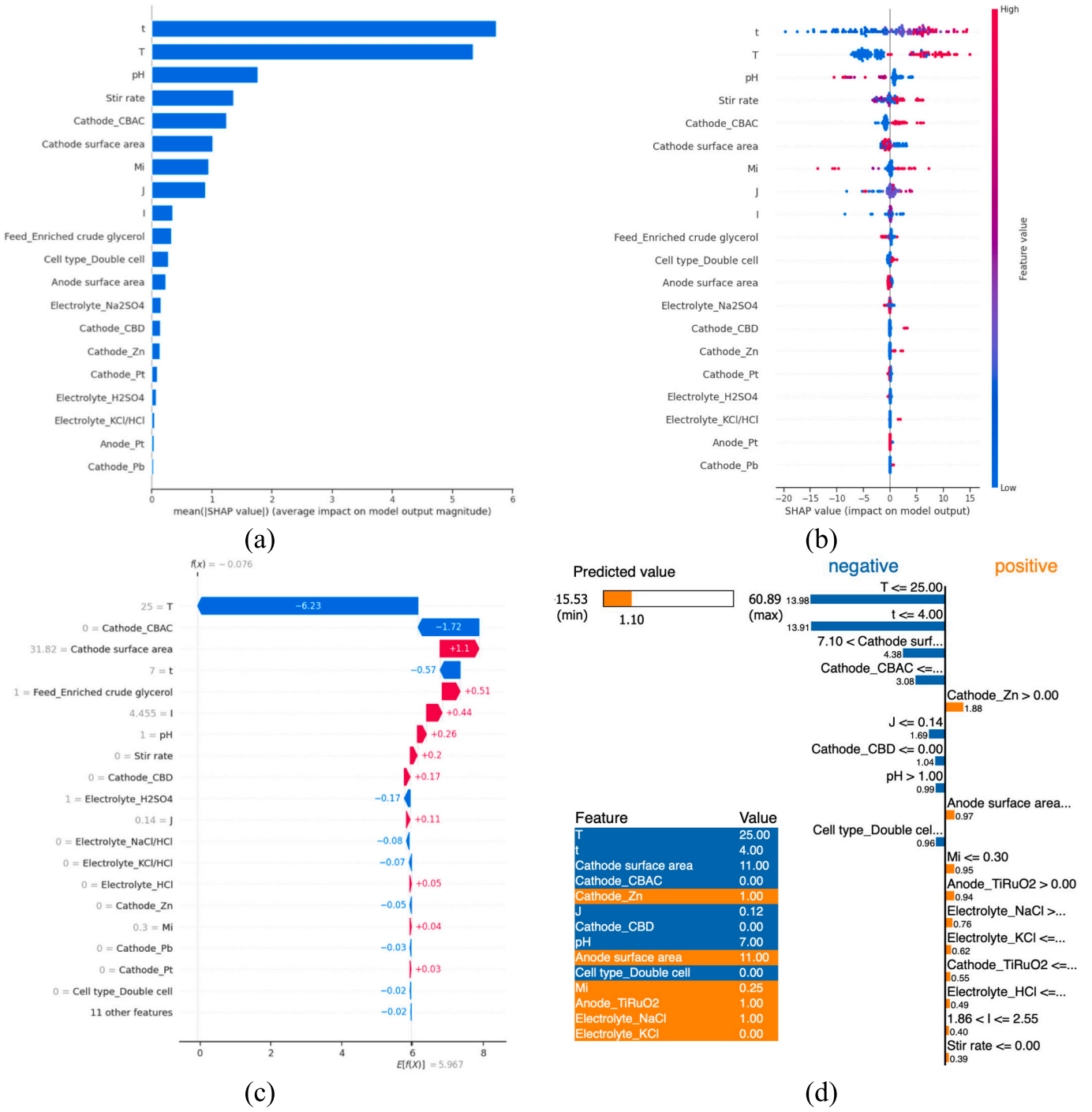

Fig. 7. Feature engineering for the XGBoost prediction model of selective ECR of glycerol to PDOs with ECR PY as the label. (a) Feature importance scores based on mean absolute SHAP values, (b) summary plot of SHAP value based on its impact on model output. (c) SHAP waterfall plots used to predict ECR PY output in our dataset. (d) LIME analysis for the prediction of ECR PY.

trained XGBoost model becomes imperative for progressing with the inverse design of the reaction.

Based on Fig. S2, the initial parameters of the PSO in this study encompassed factors such as population size, positions, velocities of particles, iterations, and constraint ranges for variables (Table S3). Assumptions were made that electrode type and properties, and the single cell reactor were constants. This allowed manipulation of reaction conditions as variables–reaction time, temperature, pH of electrolyte, stirring rate, initial glycerol concentration, current density, and current. The PSO fitness function incorporated predictions from the previously trained XGBoost model, optimizing for (i) total glycerol CR and (ii) ECR PY. In this study, the XGBoost-PSO framework was executed 10 times, generating multiple sets of optimized reaction conditions. Among these outputs, three sets of conditions were selected for experimental validation, each tested in triplicate to ensure reproducibility and statistical robustness.

3.4.1. *Optimization of glycerol CR and ECR PY*

In the first part, PSO was employed for optimizing the glycerol CR. The iterative optimization process was crucial for refining the operational conditions, ensuring robustness and reliability in the data. Conducted across multiple iterations (10 times), the PSO algorithm

consistently produced optimized CR values ranging from 95 % to 100 %, illustrated in Fig. 8(a). As in line with our feature engineering analysis, maximum CR values able to be achieved at condition of higher reaction temperature, lower pH of electrolyte as well as higher stirring rate which detailed in Table S4(a).

Furthermore, for the second part, the PSO optimization also extended to improve the ECR PY, achieving optimal values ranging from 47 % to 54 %, as documented in Table S4(b) and Fig. 8(b). These results notably surpassed previous reported experimental optimizations by Rahim et al. [15] and Hunsom and Saila [10], underscoring the efficacy of integrating ML techniques for enhanced electrocatalytic performance, despite of inconsistency of extracted data with fundamental knowledge gaps regarding the reaction. This underscores the robustness and adaptability of ML-driven optimization approaches in overcoming traditional limitations and pushing the boundaries of electrochemical efficiency [50,68].

3.4.2. Multi-objective optimization considerations

The current optimization framework primarily focuses on optimizing individual objectives of CR and ECR PY, rather than simultaneously achieving optimal values for both. From our study, it becomes evident that the optimal input parameters for CR and ECR PY have slightly differ. This is due to the fundamentally different reaction pathways and mechanisms got involve in achieving high CR values. In an electrolytic cell, high CR can be achieved either from reduction reaction and/or oxidation reaction as well as other complex side reactions [5,62]. Thus, combinining these outputs using multi-output models will result to a trade-off [69,70]. Specifically, while optimizing for CR, ECR PY often deviates from its optimal values, and vice versa. This indicates a fundamental challenge in balancing the optimization of both objectives simultaneously. Thus, a detailed study is needed to address this particular knowledge gap for further understanding. In addition, this aligns with literature observations, where conditions that favor glycerol electrooxidation reactions tend to dominate the optimization of CR, while those favorable for glycerol electroreduction align more closely with ECR PY.

In short, given the primary objective of this study is to evaluate the performance of glycerol ECR to produce PDOs, we believe focusing on individual optimizations for CR and ECR PY is sufficient to provide valuable insights. The optimal values and corresponding conditions for ECR PY, as our main focus, are detailed in Fig. 8(b). In addition, our optimization results align with the knowledge and insights gleaned from our earlier feature engineering analysis (Subsection 3.3, Figs. 6 and 7, respectively).

3.5. Experimental verification for glycerol CR and ECR PY

For experimental validation, we experimentally applied the parameter combination suggested by our XGBoost-PSO model. However, due to limitations in experimental equipment, as compared to the dataset used in the prediction model–such as different electrode types, sizes, and geometric surface areas along with the use of a lower-limit potentiostat–we adjusted the parameter boundaries within the PSO algorithm (Table S3). As a result, the new optimized CR and PY values were lower than the previous optimization. A 3D data visualization illustrating the reduction in output values due to these experimental limitations is provided in Fig. 9. The updated maximum CR value was 72.56 %, achieved under the optimized conditions of t = 9.37 h, T = 39.94 °C, pH = 2.55, Stir rate = 692.91 rpm, Mi = 0.91 M, J = 0.003 A/cm^2, and I = 0.016 A with Pt cathode (surface area 0.17 cm^2) (Fig. 9(a)). Meanwhile, the new maximum ECR PY recorded was 32.49 %, attained at t = 23.56 h, T = 80.08 °C, pH = 1, Stir rate = 259.26 rpm, Mi = 0.51 M, J = 0.010 A/cm^2, and I = 0.010 A with carbon cathode (Fig. 9(b)). The detailed optimized conditions and corresponding optimized results are presented in Tables S5(a) and S5(b).

3.5.1. Experimental vs ML-predicted outputs

To validate the PSO-optimized conditions, which presented in Table S5, we experimentally tested three sets of optimized condition combinations yielding the highest output values. Specifically, for CR, we tested conditions No. 1, No. 2, and No. 4 from Table S5(a), and for ECR PY, conditions No. 4, No. 6, and No. 8 from Table S5(b). For CR outputs, the experimentally verified value under the predicted optimal conditions (No. 1, Table S5(a)) was 61.06 %, compared to the predicted value of 70.13 %, yielding an experimental error of 9.07 %. Similar deviations were recorded for No. 2 (12.14 %) and No. 4 (9.75 %), resulting in an average CR error of 10.32 ± 1.61 % (Fig. 10(a)). Meanwhile, for ECR PY, the experimental value for No. 4 (Table S5(b)) was 23.08 %, compared to the predicted values for ECR PY was 31.33 %, with a low experimental error of 8.25 %. Errors for No. 6 (8.88 %) and No. 8 (13.87 %) led to an average ECR PY error of 10.33 ± 3.08 % (Fig. 10(b)).

The processing parameters, predicted by our XGBoost-PSO model, resulted in experimental conversion rates and product yields with average experimental errors around 10 %. These results were aligned

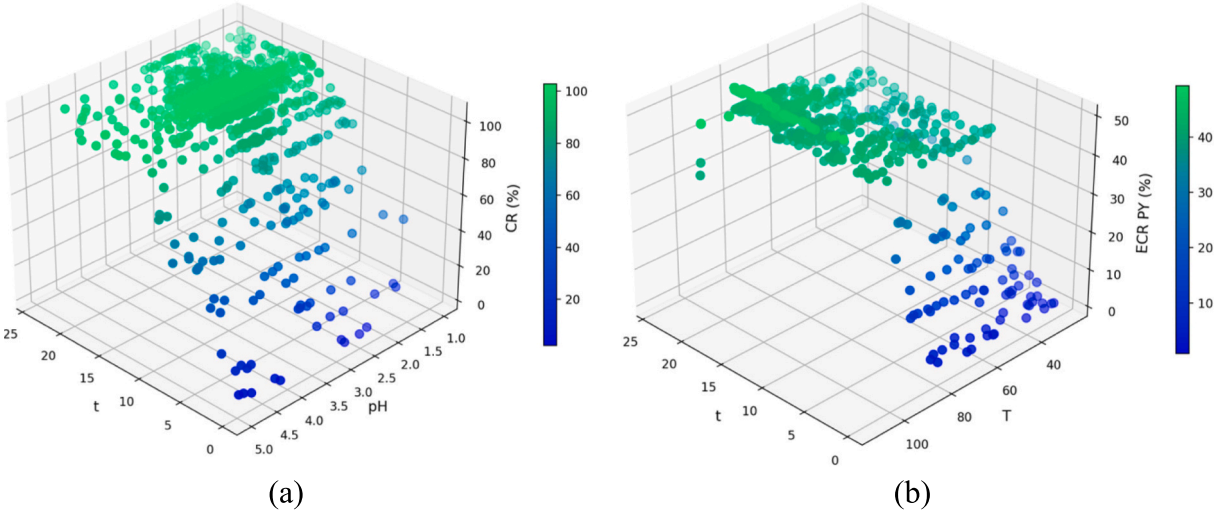

Fig. 8. Optimization results based on output variables of (a) CR and (b) ECR PY, as fitness function using PSO approach. Table S4, as reported in the supplementary details, thoroughly recorded the combination of optimal reaction conditions for CR and ECR PY, respectively, which run under no experimental equipment limitations.

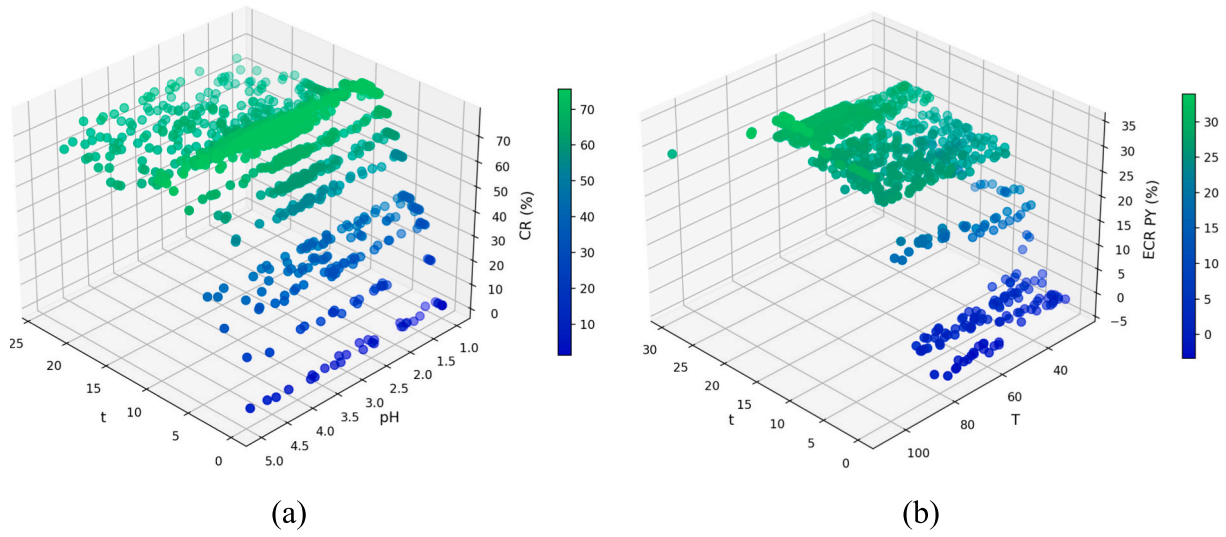

Fig. 9. Optimization results based on output variables of (a) CR and (b) ECR PY, as fitness function using PSO approach after adjusting the features' boundaries.

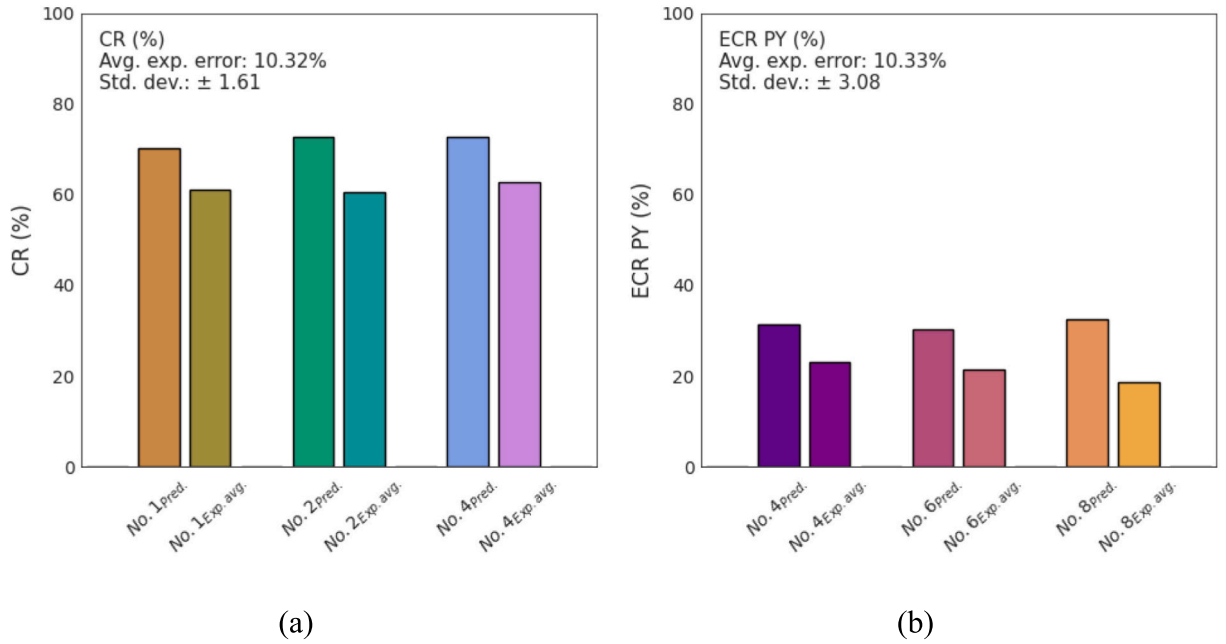

Fig. 10. Experimental verification (bar plots) for glycerol conversion rate (CR) (a) and electroreduction product yield (ECR PY) (b) under the optimized conditions derived from the well-trained XGBoost-PSO algorithm.

with literature-reported ML-based predictions of Li et al. [71], which reported calculation errors of <20 %. These errors were deemed subjectively acceptable from an engineering standpoint, which also reported by other literature [40,72,73]. To reinforce the validation of the optimized processing parameters derived from the ML models, each condition for both CR and ECR PY outputs was triplicated (Tables S5(a) and (b)), showing consistent results with a low standard deviation for the average experimental error of ±1.61 % and ± 3.08 %, stated in Figs. 10(a) and (b). In contrast to traditional experimental approaches, which typically demand numerous experimental works for optimization, the ML-based inverse design strategy significantly reduced the number of experimental runs while achieving comparable results. In this study, eight processing parameters were successfully optimized, demonstrating the effectiveness of ML-assisted inverse design in reducing the experimental workload.

3.5.2. *GC–MS analysis of electrochemical testing*

GC–MS analysis for CR experiment of optimized reaction conditions in No. 2 (Table S5(a)) was revealed in Fig. S4(b). The CR value under this condition was 60.42 %, while the ECR PY was extremely low, with no detectable formation of PDOs. Instead, GC–MS analysis identified alternative products such as propanoic acid, acetone, dimethoxybutane and methyl dioxolane. These results are in line with several published studies [11,13,74], which reported that certain combinations of reaction conditions favored electrooxidation reactions–producing propanoic acid and acetone–rather than electroreduction. Additionally, the formation of dimethoxybutane and methyl dioxolane suggests that side reactions such as acetalization also occurred under these conditions [75].

Similarly, in the subsequent part, under the optimized conditions of No. 8 in Table S5(b), we detected 1,3-PDO, 1,2-PDO (also named propylene glycol), dioxolane diol, and dihydroxyacetone as the main compounds as shown in Fig. S4(d). From Fig. 10(b), the average yield of

PDOs was recorded at approximately 21.01 ± 2.25 %. These conditions successfully optimizing electro-deoxygenation reaction which facilitating the breakdown of C—O bond within glycerol structure, consistent with the literature [11,15,16]. The CR value was also significantly higher, reaching approximately 100 %. This finding suggests that while the single-objective PSO optimization was designed to focus on ECR PY, the selected conditions also contributed positively to CR. Specifically, the longer processing time, lower electrolyte pH, higher stirring rate, and significantly elevated temperature—factors analyzed in Fig. 6—played a critical role in improving the CR value during glycerol ECR. This highlights the versatility of our approach and confirms the validity of our XGBoost-PSO prediction and feature analysis in predicting optimized combination of reaction conditions which favors the electroreduction of glycerol.

3.6. Challenges and future outlooks

The experimental verification process identified subtle experimental errors originated from various factors including both experimental procedures as well as ML modeling. From our investigation, it is mainly due to the inconsistencies in the experimental setups across studies. Factors such as power supply capabilities, cell designs and sizes, stirrer type and heating methods varied between our experimental setup and those referenced in prior literature. This inconsistency can introduce deviations, especially when specific details of these setups were not disclosed in previous studies, making exact replication challenging. This is an unavoidable issue where such variations influence reaction conditions and therefore contribute to differences between predicted and actual performance metrics.

Our equipment has certain limitations, such as a maximum allowable current flow of 0.02 A due to the potentiostat's limited ability to apply voltages exceeding 15 V, particularly in electrolytes with higher resistance. This constraint results in suboptimal reaction performance compared to the literature, where higher current densities are sometimes achievable. Despite these limitations, our PSO algorithm able to handle it by updating the optimization boundaries according to our existing limitations, proofs its practicality and adaptability. Such adaptability not only enhances the precision of optimization outcomes but also exemplifies how ML methodologies can transcend traditional experimental limitations, thereby advancing electrochemical efficiency and performance predictability. However, some discrepancies remain due to the intrinsic limitations of our experimental hardware.

Next, based on our feature importance outputs, the working electrode (electrocatalysts) properties play a crucial role in glycerol ECR, significantly influencing reaction rates and product yields. In our experiments, we used commercial carbon electrodes, which differ from the custom-fabricated carbon black/activated carbon (*Cathode_CBAC*) electrode [12], to achieve 32.49 % of ECR PY (No. 8 in Table S5(b)), recommended by the XGBoost-PSO model for optimal ECR PY yield. Different in electrode materials impact factors like their geometric surface area, electronic conductivity, and catalytic properties, all of which are crucial for selective glycerol ECR [11,76]. These material differences likely to contribute to the experimental error in ECR PY prediction, as carbon electrode may not have the same catalytic efficiency as the self-fabricated CBAC electrodes proposed by the optimization model.

While the XGBoost model performs well, the model may not fully capture nuanced variations between setups, especially given the limited dataset size. Furthermore, data limitations often necessitate generalizations that can introduce predictive inaccuracies when applied to specific experimental setups. To minimize disparities between predicted and experimental values, it is imperative to generate our own dataset under a consistent experimental setup through high-throughput computations and experiments [77] to be fed into our ML prediction model. By applying this measure, both systematic errors, such as calibration inconsistencies or measurement biases, and random errors, including environmental fluctuations or human handling, within practical experiments can be controlled as well as minimized. However, these do not diminish the utility of our ML-based optimization approach. Instead, they highlight areas for refinement, offering valuable insights for improving both experimental design and model accuracy.

On the other hand, addressing the abovementioned challenges is essential for further improving prediction accuracy, representing a significant gap in this ML-chemistry niche, are very crucial [48,63,73]. Our planned future work includes generating a high-throughput dataset under consistent experimental conditions, applying more detailed techniques like hyperparameter tuning, transfer learning, and active learning to refine our model, and constructing more comprehensive materials database. This database would integrate advanced features of electrocatalysts, such as geometric (atomic radii, coordination numbers, etc.), electronic (electronegativity value, standard reduction potential, conductivity, etc.) and catalytic properties, which could enhance the model's ability to predict outcomes accurately [37]. This integration can significantly improve the predictive accuracy. Thus, our forthcoming focus will be on optimizing electrocatalyst fabrication using ML-based optimization method as well as to consider the stability of the catalyst over extended periods via various specific techniques via cyclic voltammetry (CV) and chronopotentiometry (CP) or chronoamperometry (CA) [78], which directly affect the electrocatalytic reduction reactions to achieve maximal catalytic reactivity. Leveraging ML, we aim to not only optimize product yields but also to control and sustain the reactions happened within the system, ensuring that catalyst performance remains robust over prolonged operational times. These focuses will be the key components for our future research.

4. Conclusion

This study establishes a robust machine learning framework, by integrating XGBoost with PSO algorithm, for optimizing the selective electrocatalytic reduction of glycerol to value-added biochemicals, particularly 1,2- and 1,3-PDOs, demonstrating how data-driven approaches can overcome traditional limitations in electrochemical process development. Through systematic evaluation of six widely used ML algorithms, XGBoost emerged as the optimal model, achieving exceptional predictive performance for both CR and ECR PY outputs with R^2_{test} scores of 0.98 and 0.80, respectively, while maintaining interpretability through SHAP and LIME analysis–a critical advantage for guiding experimental validation. This model outperforms other models such as ANN and GPR and proving its robustness in handling imbalanced datasets. Additionally, the integration of PSO yielded theoretically optimal conditions achieving 100.26 % conversion rate and 53.29 % PDO yield at consistent combination of reaction conditions, representing a 25–40 % improvement over conventional trial-and-error optimization approaches reported in literature. Experimental verification of three optimized combination of conditions per output showed high reliability and significant prediction accuracy, with an average experimental error of 10.32 ± 1.61 % for CR and 10.33 ± 3.08 % for ECR PY.

Throughout the study, three key mechanistic insights emerged from our analysis:

1. Process condition synergies. We identified alongside to longer processing time, with acidic conditions (pH 1–3) and use of Pt cathode favor high CR, whereas, interestingly, high temperatures (80–100 °C) coupled with carbon-based electrocatalyst preferentially facilitating C—O bond cleavage for selective PDO formation. These findings revealing previously underappreciated thermodynamic-kinetic tradeoffs in glycerol ECR pathways.
2. Time-dependent selectivity. SHAP and LIME analysis demonstrated reaction time as the dominant feature for both ECR PY and CR (with the highest SHAP and LIME values), but with divergent optimal ranges, highlighting the need for staged process conditions to maximize overall efficiency.

3. ML-driven optimization limits. Experimental validation confirmed the XGBoost-PSO model's predictive capability (~10.3 % experimental error) while revealing equipment-specific boundary conditions that must be incorporated for industrial translation–particularly regarding current density limitations and electrocatalyst fabrication variability.

The success of our XGBoost-PSO framework suggests three transformative opportunities for the field, which includes:

1. Advanced electrocatalyst discovery with the implementation of GANs for inverse design of tailored non-noble metallic electrocatalysts that simultaneously address catalytic activity, selectivity and stability requirements.
2. Standardized high-throughput experimental validation, generating the homogenized datasets needed to address current reproducibility challenges in literature.
3. Multi-objective process optimization. Development of constrained ML architectures to balance competing objectives–conversion rate, targeted product yield, and energy efficiency–while incorporating real-world equipment constraints.
4. Implementation of active learning for process optimization. Creation of federated learning systems that continuously incorporate new experimental data to overcome current data scarcity limitations.

From a practical standpoint, this data-driven approach is crucial for addressing knowledge gaps in the literature relating to the valorization of crude glycerol where indirectly provides both immediate and long-term value to the biodiesel industry. The validated models can already reduce experimental optimization workloads, accelerating the systematic development of selective glycerol ECR applications while significantly enhancing the economic viability of the biodiesel industry, aligned with circular bioeconomy principles.

CRediT authorship contribution statement

Muhammad Harussani Moklis: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Cries Avian:** Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Formal analysis. **Cheng Shuo:** Writing – review & editing, Validation, Supervision, Resources, Funding acquisition. **Sasipa Boonyubol:** Writing – review & editing, Validation, Supervision, Investigation, Formal analysis. **Jeffrey S. Cross:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

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.

Acknowledgements

The authors express their gratitude to the Ministry of Education, Culture, Sports, Science, and Technology (Monbukagakusho): MEXT scholarship and Institute of Science Tokyo (formerly known as Tokyo Institute of Technology) for their financial assistance and support in conducting this research. Additionally, special thanks are extended to Dr. Abraham C. Garcia for his invaluable contribution in research methodology and guidance for sample analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jelechem.2025.119150.

Data availability statement

The codes of the models and the collected datasets are available at https://github.com/MMHarussani/respaper1-glycerolECRMLPSO. Data will be made available on request.

References

[1] International Energy Agency (IEA), Biofuels, IEA, France. https://www.iea.org/reports/biofuels, 2022.
[2] M. Athar, S. Zaidi, A review of the feedstocks, catalysts, and intensification techniques for sustainable biodiesel production, J. Environ. Chem. Eng. 8 (2020) 104523.
[3] M.H. Moklis, S. Cheng, J.S. Cross, Current and future trends for crude glycerol upgrading to high value-added products, Sustainability 15 (2023) 2979.
[4] S. Jaiswal, S. Sahani, Y.C. Shar, Enviro-benign synthesis of glycerol carbonate utilizing bio-waste glycerol over Na-Ti based heterogeneous catalyst: kinetics and E-metrics studies, J. Environ. Chem. Eng. 10 (2022) 107485.
[5] M.H. Moklis, C. Shuo, S. Boonyubol, J.S. Cross, Electrochemical valorization of glycerol via Electrocatalytic reduction into biofuels: a review, ChemSusChem 17 (2023) e202300990.
[6] A.K. Mostafazadeh, M.S. De La Torre, Y. Padilla, P. Drogui, S.K. Brar, R.D. Tyagi, Y. Le Bihan, G. Buelna, P.G. Moroyoqui, An insight into an electro-catalytic reactor concept for high value-added production from crude glycerol: optimization, electrode passivation, product distribution, and reaction pathway identification, Renew. Energy 172 (2021) 130–144.
[7] C. Schotten, T.P. Nicholls, R.A. Bourne, N. Kapur, B.N. Nguyen, C.E. Willans, Making electrochemistry easily accessible to the synthetic chemist, Green Chem. 22 (2020) 3358–3375.
[8] U. Sanyal, J. Lopez-Ruiz, A.B. Padmaperuma, J. Holladay, O.Y. Gutiérrez, Electrocatalytic hydrogenation of oxygenated compounds in aqueous phase, Org. Process. Res. Dev. 22 (2018) 1590–1598.
[9] J.A. Lopez-Ruiz, E. Andrews, S.A. Akhade, M.-S. Lee, K. Koh, U. Sanyal, S.F. Yuk, A. J. Karkamkar, M.A. Derewinski, J. Holladay, Understanding the role of metal and molecular structure on the electrocatalytic hydrogenation of oxygenated organic compounds, ACS Catal. 9 (2019) 9964–9972.
[10] M. Hunsom, P. Saila, Electrochemical conversion of enriched crude glycerol: effect of operating parameters, Renew. Energy 74 (2015) 227–236.
[11] O.O. James, W. Sauter, U. Schröder, Towards selective electrochemical conversion of glycerol to 1, 3-propanediol, RSC Adv. 8 (2018) 10818–10827.
[12] C.S. Lee, M.K. Aroua, W.M.A.W. Daud, P. Cognet, Y. Peres-Lucchese, M.A. Ajeel, Selective electroreduction of glycerol to 1, 2-propanediol on a mixed carbon-black activated carbon electrode and a mixed carbon black-diamond electrode, BioResources 13 (2018) 1–16.
[13] Z. Liang, M.A. Villalba, G. Marcandalli, K. Ojha, A.J. Shih, M.T. Koper, Electrochemical reduction of the simplest monosaccharides: dihydroxyacetone and glyceraldehyde, ACS Catal. 10 (2020) 13895–13903.
[14] E. Peralta-Reyes, D. Vizarretea-Vásquez, R. Natividad, A. Aizpuru, E. Robles-Gómez, C. Alanis, A. Regalado-Mendez, Electrochemical reforming of glycerol into hydrogen in a batch-stirred electrochemical tank reactor equipped with stainless steel electrodes: parametric optimization, total operating cost, and life cycle assessment, J. Environ. Chem. Eng. 10 (2022) 108108.
[15] S.A.N.M. Rahim, C.S. Lee, F. Abnisa, W.M.A.W. Daud, M.K. Aroua, P. Cognet, Y. Pérès, Activated carbon-based electrodes for two-steps catalytic/electrocatalytic reduction of glycerol in Amberlyst-15 mediator, Chemosphere 295 (2022) 133949.
[16] W. Sauter, O.L. Bergmann, U. Schröder, Hydroxyacetone: a glycerol-based platform for Electrocatalytic hydrogenation and Hydrodeoxygenation processes, ChemSusChem 10 (2017) 3105–3110.
[17] S. Verma, S. Lu, P.J. Kenis, Co-electrolysis of CO2 and glycerol as a pathway to carbon chemicals with improved technoeconomics due to low electricity consumption, Nat. Energy 4 (2019) 466–474.
[18] E. Fontes, C. Ramos, C. Ottoni, R. de Souza, E. Antolini, A. Neto, Glycerol dehydrogenation steps on au/C surface in alkaline medium: an in-situ ATR-FTIR approach, Renew. Energy 167 (2021) 954–959.
[19] M. Nacef, M.L. Chelaghmia, O. Khelifi, M. Pontié, M. Djelaibia, R. Guerfa, V. Bertagna, C. Vautrin-Ul, A. Fares, A.M. Affoune, Electrodeposited Ni on pencil graphite electrode for glycerol electrooxidation in alkaline media, Int. J. Hydrog. Energy 46 (2021) 37670–37678.
[20] I. Velázquez-Hernández, E. Zamudio, F.J. Rodríguez-Valadez, N.A. García-Gómez, L. Álvarez-Contreras, M. Guerra-Balcázar, N. Arjona, Electrochemical valorization of crude glycerol in alkaline medium for energy conversion using Pd, Au and PdAu nanomaterials, Fuel 262 (2020) 116556.
[21] M.S. Houache, K. Hughes, E.A. Baranova, Study on catalyst selection for electrochemical valorization of glycerol, sustain, Energy Fuel 3 (2019) 1892–1915.
[22] J. Iyyappan, B. Bharathiraja, G. Baskar, E. Kamalanaban, Process optimization and kinetic analysis of malic acid production from crude glycerol using aspergillus Niger, Bioresour. Technol. 281 (2019) 18–25.
[23] C. Chen, G.X. Gu, Generative deep neural networks for inverse materials design using backpropagation and active learning, Adv. Sci. 7 (2020) 1902607.

[24] A. Mistry, A.A. Franco, S.J. Cooper, S.A. Roberts, V. Viswanathan, How machine learning will revolutionize electrochemical sciences, ACS Energy Lett. 6 (2021) 1422–1431.
[25] J. Sun, C. Liu, What and how can machine learning help to decipher mechanisms in molecular electrochemistry? Curr. Opin. Electrochem. 39 (2023) 101306.
[26] A.F. Zahrt, Y. Mo, K.Y. Nandiwale, R. Shprints, E. Heid, K.F. Jensen, Machine-learning-guided discovery of electrochemical reactions, J. Am. Chem. Soc. 144 (2022) 22599–22610.
[27] K. Karunamurthy, A.A. Janvekar, P. Palaniappan, V. Adhitya, T. Lokeswar, J. Harish, Prediction of IC engine performance and emission parameters using machine learning: a review, J. Therm. Anal. Calorim. 148 (2023) 3155–3177.
[28] H. Aygun, O.O. Dursun, K. Dönmez, O. Sahin, S. Toraman, Prediction of performance characteristics of an experimental micro turbojet engine using machine learning approaches, Energy 313 (2024) 133997.
[29] X. Yang, M. Li, J. Shen, Z. Liu, W. Liu, R. Long, Deep learning assisted anode porous transport layer inverse design for proton exchange membrane water electrolysis, Int. J. Heat Mass Transf. 233 (2024) 126019.
[30] X. Chen, A. Rex, J. Woelke, C. Eckert, B. Bensmann, R. Hanke-Rauschenbach, P. Geyer, Machine learning in proton exchange membrane water electrolysis—a knowledge-integrated framework, Appl. Energy 371 (2024) 123550.
[31] S.N. Ozdemir, O. Pektezel, Performance prediction of experimental PEM electrolyzer using machine learning algorithms, Fuel 378 (2024) 132853.
[32] K. Kumar, B. Pande, Air pollution prediction with machine learning: a case study of Indian cities, Int. J. Environ. Sci. Technol. 20 (2023) 5333–5348.
[33] M. Hunsom, P. Saila, Product distribution of electrochemical conversion of glycerol via Pt electrode: effect of initial pH, Int. J. Electrochem. Sci. 8 (2013) 11288–11300.
[34] S. Kongjao, S. Damronglerd, M. Hunsom, Electrochemical reforming of an acidic aqueous glycerol solution on Pt electrodes, J. Appl. Electrochem. 41 (2011) 215–222.
[35] S.A.N.M. Rahim, C.S. Lee, F. Abnisa, M.K. Aroua, W.A.W. Daud, P. Cognet, Y. Pérès, A review of recent developments on kinetics parameters for glycerol electrochemical conversion–a by-product of biodiesel, Sci. Total Environ. 705 (2020) 135137.
[36] M.K. Dahouda, I. Joe, A deep-learned embedding technique for categorical features encoding, IEEE Access 9 (2021) 114381–114391.
[37] M.R. Dobbelaere, P.P. Plehiers, R. Van de Vijver, C.V. Stevens, K.M. Van Geem, Machine learning in chemical engineering: strengths, weaknesses, opportunities, and threats, Engineering 7 (2021) 1201–1211.
[38] A. Adadi, M. Berrada, Peeking inside the black-box: a survey on explainable artificial intelligence (XAI), IEEE Access 6 (2018) 52138–52160.
[39] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat. Mach. Intell. 1 (2019) 206–215.
[40] Y. Sun, Z. Zhao, H. Tong, B. Sun, Y. Liu, N. Ren, S. You, Machine learning models for inverse design of the electrochemical oxidation process for water purification, Environ. Sci. Technol. 57 (2023) 17990–18000.
[41] C. Xiao, J. Ye, R.M. Esteves, C. Rong, Using spearman's correlation coefficients for exploratory data analysis on big dataset, Concurr. Comput. Pract. Exp. 28 (2016) 3866–3878.
[42] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, Explainable ai: a review of machine learning interpretability methods, Entropy 23 (2020) 18.
[43] L. Cheng, S. Ramchandran, T. Vatanen, N. Lietzén, R. Lahesmaa, A. Vehtari, H. Lähdesmäki, An additive Gaussian process regression model for interpretable non-parametric analysis of longitudinal data, Nat. Commun. 10 (2019) 1798.
[44] R. Pan, M. Gu, J. Wu, Physics-informed Gaussian process regression of in operando capacitance for carbon supercapacitors, Energy Adv. 2 (2023) 843–853.
[45] J.L. Hitt, D. Yoon, J.R. Shallenberger, D.A. Muller, T.E. Mallouk, High-throughput fluorescent screening and machine learning for feature selection of Electrocatalysts for the alkaline hydrogen oxidation reaction, ACS Sustain. Chem. Eng. 10 (2022) 16299–16312.
[46] B. Azhar, C. Avian, A.H. Tiwikrama, Green synthesis optimization with artificial intelligence studies of copper–gallic acid metal–organic framework and its application in dye removal from wastewater, J. Mol. Liq. 389 (2023) 122844.
[47] M. Shariq, S. Marimuthu, A.R. Dixit, S. Chattopadhyaya, S. Pandiaraj, M. Muthuramamoorthy, A.N. Alodhyab, M.K. Nazeeruddin, A.N. Grace, Machine learning models for prediction of electrochemical properties in supercapacitor electrodes using MXene and graphene nanoplatelets, Chem. Eng. J. 484 (2024) 149502.
[48] J.A. Esterhuizen, B.R. Goldsmith, S. Linic, Interpretable machine learning for knowledge generation in heterogeneous catalysis, Nat. Catal. 5 (2022) 175–184.
[49] O. Sagi, L. Rokach, Approximating XGBoost with an interpretable decision tree, Inf. Sci. 572 (2021) 522–542.
[50] R. Ding, J. Chen, Y. Chen, J. Liu, Y. Bando, X. Wang, Unlocking the potential: machine learning applications in electrocatalyst design for electrochemical hydrogen energy transformation, Chem. Soc. Rev. 53 (2024) 11390–11461.
[51] A. Zaffar, M.R. Prabhakar, C. Liu, J. Sivaraman, P. Balasubramanian, Machine learning assisted prediction and process validation of electrochemically induced phosphorus recovery from wastewater, J. Environ. Chem. Eng. 12 (2024) 114271.
[52] C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, A comparative analysis of gradient boosting algorithms, Artif. Intell. Rev. 54 (2021) 1937–1967.
[53] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (2009) 1263–1284.
[54] H. Song, M. Kim, D. Park, Y. Shin, J.-G. Lee, Learning from noisy labels with deep neural networks: a survey, IEEE Trans. Neural Netw. Learn. Syst. 34 (2022) 8135–8153.
[55] S.B.S. Lai, N. Shahri, M.B. Mohamad, H. Rahman, A.B. Rambli, Comparing the performance of AdaBoost, XGBoost, and logistic regression for imbalanced data, Math. Stat. 9 (2021) 379–385.
[56] F. Xiao, T. Chen, J. Zhang, S. Zhang, Water management fault diagnosis for proton-exchange membrane fuel cells based on deep learning methods, Int. J. Hydrog. Energy 48 (2023) 28163–28173.
[57] X. Yuan, M. Suvarna, S. Low, P.D. Dissanayake, K.B. Lee, J. Li, X. Wang, Y.S. Ok, Applied machine learning for prediction of CO2 adsorption on biomass waste-derived porous carbons, Environ. Sci. Technol. 55 (2021) 11925–11936.
[58] A. Castro Garcia, S. Cheng, S.E. McGlynn, J.S. Cross, Machine learning model insights into base-catalyzed hydrothermal lignin Depolymerization, ACS Omega 8 (2023) 32078–32089.
[59] J. Park, J. Lee, Gradient boosting algorithm for current-voltage prediction of fuel cells, Electrochim. Acta 432 (2022) 141148.
[60] B. Gregorutti, B. Michel, P. Saint-Pierre, Correlation and variable importance in random forests, Stat. Comput. 27 (2017) 659–678.
[61] S.D. Blass, R.J. Hermann, N.E. Persson, A. Bhan, L.D. Schmidt, Conversion of glycerol to light olefins and gasoline precursors, Appl. Catal. A Gen. 475 (2014) 10–15.
[62] M. Simões, S. Baranton, C. Coutanceau, Electrochemical valorisation of glycerol, ChemSusChem 5 (2012) 2106–2124.
[63] J. Liu, W. Luo, L. Wang, J. Zhang, X. Fu, J. Luo, Toward excellence of electrocatalyst design by emerging descriptor-oriented machine learning, Adv. Funct. Mater. 32 (2022) 2110748.
[64] M.A. Kishore, S. Lee, J.S. Yoo, Fundamental limitation in electrochemical methane oxidation to alcohol: a review and theoretical perspective on overcoming it, Adv. Sci. 10 (2023) 2301912.
[65] T. Đukić, L.J. Moriau, L. Pavko, M. Kostelec, M. Prokop, F. Ruiz-Zepeda, M. Šala, G. Dražić, M. Gatalo, N. Hodnik, Understanding the crucial significance of the temperature and potential window on the stability of carbon supported Pt-alloy nanoparticles as oxygen reduction reaction electrocatalysts, ACS Catal. 12 (2021) 101–115.
[66] L. Liu, A. Corma, Structural transformations of solid electrocatalysts and photocatalysts, Nat. Rev. Chem. 5 (2021) 256–276.
[67] F. Zeng, C. Mebrahtu, L. Liao, A.K. Beine, R. Palkovits, Stability and deactivation of OER electrocatalysts: a review, J. Energy Chem. 69 (2022) 301–329.
[68] J. Park, J. Lee, Electrochemical energy conversion and storage processes with machine learning, Trends Chem 6 (2024) 302–313.
[69] Z. Li, J. Yu, C. Wang, I.T. Bello, N. Yu, X. Chen, K. Zheng, M. Han, M. Ni, Multi-objective optimization of protonic ceramic electrolysis cells based on a deep neural network surrogate model, Appl. Energy 365 (2024) 123236.
[70] G. Zhang, Z. Yan, Q. Liu, G. Lu, Multi-objective optimization strategy for multipolar magnesium electrolysis cell based on thermal-electric model, Chem. Eng. J. 490 (2024) 151690.
[71] J. Li, L. Zhang, C. Li, H. Tian, J. Ning, J. Zhang, Y.W. Tong, X. Wang, Data-driven based in-depth interpretation and inverse design of anaerobic digestion for CH4-rich biogas production, ACS EST Eng. 2 (2022) 642–652.
[72] P.H. Amaral, A.D. Torrez-Baptista, D. Dionisio, T. Lopes, J.R. Meneghini, C. R. Miranda, A machine learning model for adsorption energies of chemical species applied to CO2 Electroreduction, J. Electrochem. Soc. 169 (2022) 116505.
[73] Y. Guo, X. He, Y. Su, Y. Dai, M. Xie, S. Yang, J. Chen, K. Wang, D. Zhou, C. Wang, Machine-learning-guided discovery and optimization of additives in preparing cu catalysts for CO2 reduction, J. Am. Chem. Soc. 143 (2021) 5755–5762.
[74] Y. Nakagawa, M. Tamura, K. Tomishige, Catalytic materials for the hydrogenolysis of glycerol to 1, 3-propanediol, J. Mater. Chem. A 2 (2014) 6688–6702.
[75] M. Gonçalves, R. Rodrigues, T.S. Galhardo, W.A. Carvalho, Highly selective acetalization of glycerol with acetone to solketal over acidic carbon-based catalysts from biodiesel waste, Fuel 181 (2016) 46–54.
[76] N.K. Mishra, P. Kumar, V.C. Srivastava, U.L. Štangar, Synthesis of cu-based catalysts for hydrogenolysis of glycerol to 1, 2-propanediol with in-situ generated hydrogen, J. Environ. Chem. Eng. 9 (2021) 105263.
[77] P. Xu, X. Ji, M. Li, W. Lu, Small data machine learning in materials science, npj Comput. Mater. 9 (2023) 42.
[78] W. Zhai, Y. Ma, D. Chen, J.C. Ho, Z. Dai, Y. Qu, Recent progress on the long-term stability of hydrogen evolution reaction electrocatalysts, InfoMat 4 (2022) e12357.