A Data-driven Framework for the Accelerated Discovery of CO_2 Reduction Electrocatalysts

Ali Malek[1], Qianpu Wang[1], Stefan Baumann[2], Olivier Guillon[2,3], Michael Eikerling[3,4,5], Kourosh Malek[1]†

[1] NRC Energy, Mining and Environemtal(EME), 4250 Wesbrook Mall, Vancouver, V6T 1W5, Canada

[2] Institute of Energy and Climate Research, IEK-1: Materials sythesis and processing, Forschungszentrum Jülich, 52425 Jülich, Germany

[3] Jülich Aachen Research Alliance: JARA-Energy, 52425 Jülich, Germany

[4] Institute of Energy and Climate Research, IEK-13: Theory and Computation of Energy Materials, Forschungszentrum Jülich, 52425 Jülich, Germany

[5] Chair of Theory and Computation of Energy Materials, Division of Materials Science and Engineering, RWTH Aachen University, Intzestraße 5, 52072 Aachen, Germany

*** Correspondence:**

Corresponding Author

koroush.malek@nrc-cnrc.gc.ca

Keywords: CO_2 reduction, High and low temperature, Machine learning, Artificial intelligence, Materials discovery, Data analytics, Classification

Abstract

Searching for next-generation electrocatalyst materials for electrochemical energy technologies is a time-consuming and expensive process, even if it is enabled by high-throughput experimentation and large-scale first-principle calculations. In particular, the development of more active, selective and stable electrocatalysts for the CO_2 reduction reaction remains tedious and challenging. Here, we introduce a material recommendation and screening framework, and demonstrate its capabilities for certain classes of electrocatalyst materials for low or high-temperature CO_2 reduction. The framework utilizes high-level technical targets, advanced data extraction, and categorization paths, and it recommends the most viable materials identified using data analytics and property-matching algorithms. Results reveal relevant correlations govern catalyst performance under low and high-temperature conditions.

1. INTRODUCTION

CO_2 emissions are the main cause of human-made global warming. [1] To avert the direst consequences of this global change, the Paris Agreement calls for a net 80 to 95 % reduction of CO_2 emissions by 2050.[2] The rapid development of sustainable energy sources and environmentally benign storage and conversion technologies is thus a foremost goal in scientific research and technology development pursued collectively by countries around the world.

CO_2 can be used as a renewable feedstock for the production of synthetic fuels or fuel precursors such as CO, CH_3OH, and CH_4, addressing the problem of the intermittency of renewably generated energy from wind turbines and solar cells.[3, 4,5-7] This energy storage pathway renders the CO_2 reduction reaction (CO_2RR) a crucial and extensively researched electrochemical process.[8, 9]

CO_2RR processes inside an electrochemical cell require stable, cost-effective and highly performing electrocatalyst materials. The challenge of optimizing catalytic materials, electrodes and devices for the CO_2RR, calls for further investigation into factors that control their catalytic activity and stability. The electrocatalytic media, which are usually heterogeneous composites of the active material embedded into a host medium with suitable transport properties for gaseous reactants, liquid products, dissolved ions, and electrons, can undergo significant changes in structure and composition under operation through various phenomena such as Ostwald ripening, particle detachment or coagulation in nanoparticle-based catalyst layers; surface reconstruction, oxidation or passivation by irreversible adsorption; or electrolyte disintegration. Besides, inhibited mass transport due to non-optimal wetting of the porous electrode could cause additional voltage loss or limit the current density (CD) that a cell could attain.

A recently performed cost-benefit analysis has shown that electrochemical CO_2 conversion processes need to be economically viable at the system level, while the faradaic efficiency (FE) and energy efficiency (EF) must be maximized at the component and cell levels. [8, 10] The hydrogen evolution reaction (HER) is parallel process in CO_2 reduction cells, which impacts the yield of synthetic fuel or fuel precursor production.[11] Minimization of hydrogen production

requires electrocatalyst materials that are highly selective in terms of the reaction pathway to support.

The integration, testing and qualification of new catalyst materials is a tedious and time-consuming process as there are limitations even for the best catalysts due to specific compatibilities that are required with other components in a membrane electrode assembly (MEA), single cell or stack of the electrochemical device. Challenges in this context involve reactant and product transport as well as water and heat balances. These phenomena are coupled across multiple components and interfaces in a cell, and they determine 3D distributions of local reaction conditions in active electrode media. Assessing the impact of a new catalyst material on cell performance at cell and stack levels is thus a complex undertaking. An electrocatalyst material may show markedly improved activity and selectivity in a well-defined lab set-up under precisely controlled reaction conditions; but this improvement may not survive when the material is incorporated in a real cell and tested under realistic conditions.

Complex electrocatalytic media cannot be studied solely with computational studies based on density functional theory (DFT). Usually, the complexity of materials, components and physicochemical phenomena to be considered as well as the interplay of solvation effects, charge transfer and electrifc field effects at the interface, warrant a well-devised hierarchical framework in modeling and simulation that interweaves computational approaches, including DFT as well as classical simulations, microkinetic modeling of reaction mechanisms, interface and charge transfer theory, and continuum modeling of transport processes at the electrode scale, to rationalize local reaction conditions, decipher reaction mechanisms and calculate reaction rates. Considering all of these aspects, the theory-driven approach towards the development of highly active, selective and stable electrocatalysts for the CO_2RR remains a highly challenging task.[13, 14]

The discovery and scale-up of integrated materials, i.e., those materials that are integrated into a component, cell, and device to fulfill certain functionalities at the device level, require significant capacities for characterization, testing, and optimization at all structural levels. The discovery-to-demonstration pipeline of new electrocatalyst materials, including fabrication scale-up and integration with the electrolysis cell components is thus more complex to follow through than it is for simpler, so-called “molecular materials”, where *minimum* integration and

optimization is required beyond materials properties. [13,14] Apart from performance metrics related to activity, yield and selectivity, the degradation of cell components, overall system durability and overall cell lifetime present essential issues to be addressed, which are related to the stability of a catalyst material for given environmental conditions and operating regimes.

The key attributes of a successful design of CO_2 reduction cells include high mass activity of electrocatalysts to provide a low overpotential at reasonable materials cost, catalyst layer microstructure to facilitate charge and mass transfer, well-attuned wettabilities of porous transport media to optimize the water distribution across the cell, and mechanical and chemical durability. New approaches in materials design and integration are needed to realize the selective transformation of CO_2 into desired products in scale-up pilot or industrial setups.

Tremendous investigations have recently been made to design, synthesize and develop new CO_2RR electrocatalysts.[4,12] Machine learning (ML) and data-driven methods provide a powerful set of methods and tools to accelerate materials discovery.[15, 16] Fundamentally, ML is the practice of using statistical algorithms to parse data, learn from a set of indicators (performance metrics) and then make a fast determination or a prediction of target performance properties of any new data sets. ML in materials science is mostly concerned with supervised learning. One must realize that the selection of high-quality (accurate) datasets in addition to an appropriate set of descriptors is more important than the selection of the ML algorithm itself. The former would be considered as the first step for building any ML application; confirming the fact that an accurate ML is likely impossible without an accurate dataset. The suitable ML model, denoted as classification, regression, or rank ordering models, depends on the type of the desired outcome.[13]

However, describing all the complexities of the electrochemical interface within the DFT model, with respect to the number and the type of components (catalyst, solvent molecules, solvent ions, etc.), as well as the physics and chemical implications (electric fields, solvation, free energy, charge transfer, etc.), is challenging due to computational limitations. Classification models are designed to allocate a substance to a given number of categories such as *active* and *inactive catalysts*; they can be used to separate groups of molecules or materials according to the presence or absence of a target property. For instance, CO2RR electrocatalysts can be classified based on their Faradaic enfficiency or selectivity for a given product. In this

context, several statistical tools, in particular, regression models attempt to determine a function that can represent a continuous hypersurface that relates indicator variance to observable electrocatalytic properties. Regression models are used for where prediction and discovery of a missing physico-chemical property such as performance or selectivity are needed.[17] Ranking models put out the order of electrocatalysts for a specific property; they are highly useful for electrocatalysts design and discovery where the priority of one property over another is more important than its exact value.[18-20]

Recent self-learning algorithms have greatly influenced heterogeneous catalysis research due to the availability of ML analysis tools, e.g., Python Scikit-learn, TensorFlow and workflow management tools such as ASE[21], Atomate[22], and the proliferation of large public materials databases, including Materials Project[23], Novel Materials Discovery Laboratory[24], Citrination[25], CatApp[26], and AiiDA[27] and advancement of applied statistical algorithms and models.

ML models have been utilized in a variety of energy material applications to design and discover novel electrocatalyst materials with superior performance (e.g., higher energy density and higher energy conversion efficiency).[28, 29] Such models can have a transformative impact on the urgent needs for a variety of low cost CO2RR catalysts with high product selectivity and maximal performance.[18, 30-33] For instance, ML models have been used to disentangle the catalyst-adsorbate interactions for various reactions, including CO_2RR. [34, 35] A combination of advanced optimization tools based on ML and other conventional approaches was developed to predict electrocatalyst performance for CO_2 reduction and H_2 evolution.[34]

In this work, we demonstrate a data-driven framework for materials screening, which is particularly applied to low and high temperate catalysts for CO_2 reduction.[36, 37,10, 38]
A viable electrocatalyst for the CO_2RR must satisfy performance metrics related to current density, faradaic efficiency, energy efficiency, overpotential, production rate, and chemical stability. Correlations among these performance metrics at low or high temperature remain largely unknown and require extensive data analytics.

Our data-driven methodology is designed with the objectiveof integrating domain-specific data sources in order to eliminate difficulties in data collection and interpretation from multiple

sources and data types. The integration process consists of a combination of “modular” sub-processes to build "standardized energy materials data" in real-time with advanced filtering, scale-up and cognitive insights, ML, and fundamental data analytics functionalities, including visualization and big-data management tools. The recommendation system and decision module utilize high-level technical targets as input data, which can be displayed in the form of radar (or spider) charts, advanced data extraction and categorization using deep learning techniques, property-matching algorithms to search for the best viable materials that satisfy selected high-level technical targets, and finally a multi-parameter optimization to recommend top choices in connection with ML algorithms.

2. METHODOLOGY

2.1 Application-Driven Architecture

In order to offer scale-bridging capabilities to connect crucial steps in materials design-to-device integration, an application-driven architecture has been introduced and demonstrated [36]. The integral part of this architecture is an embedded master data lake, consisting of large-scale metadata for electrocatalyst materials, which is collected from various types and sources of materials data. Key technical targets such as activity (i.e., the faradaic efficiency), stability, and selectivity are generally defined at cell and device level and may also correlate differently at low or high temperatures with physicochemical properties of electrocatalyst and cell or device operating conditions. [39, 40]

Figure 1 illustrates the functional layers of the ML-enabled data analytics approach and its underlying workflow. The workflow comprises various layers including user-defined or default data sources and databases, analytics modules, and self-driving algorithms, which are generally used in any materials discovery, regardless of the corresponding field of application. The complexity with scale-up and discovery of integrated materials also implies the need for ad-hoc communication among parallel or series of synthesis and characterization steps or equipment, in-device component integration, and device testing or validation. This all-embracing workflow along the complete development pipeline can potentially enhance data communication and understanding correlations among structure, functional properties, and performance indicators at all scales from materials discovery to device performance and optimization.

comprising various data sources and physico-chemical processes, which are used in materials discovery. The main distinction is between autonomous and de-centralized approaches. For the autonomous approach, the entire precursor preparation, mixing, testing, and characterization processes are performed by an automated robotic equipment. In contrast, the de-centralized approach utilizes existing legacy equipment by employing advances in AIand the Internet of Things (IoT) connectivity. This enforces communication among different processes and equipment can take place seamlessly via cloud computing. A cognitive processwith accurate and distinct correlation functions between structure, functional properties, and performance can enhance the de-centralized approach to materials discovery. The decentralized approach can bring about a robust and rapid implementation in a more cost-effective fashion than that under an autonomous process.

2.2 Master Data Lakes

A vital prerequisite for any form of ML application is the provision of a suitable dataset for a given domain. The search for new electrocatalyst materials essentially needs a minimal and sufficient set of performance indicators from the "chemical domain" and the "property domain" of different electrocatalyst materials. [44]

The master database is formed from materials datasets collected from a wide range of sources and user-types, namely 1) unpublished records of academic researchers, 2) published articles, and 3) other public records and industry collaborators. The details of the data retrieval from images, tables and texts are described in ref [36].The resulting database is stored in excel or CSV format with predefined and standardized headers that include metadata preprocessing and cleaning.

In this article, the CO_2RR experimental databases were generated from literature sources on the basis of seven input variables; electrocatalyst type, faradaic efficiency, applied potential, current density, type of electrolyte, the major product, and temperature. Each experimental data point is characterized by a set of performance indicators for catalyst formulation and reaction conditions, either as continuous values (such as current density) or as categorical values (such as catalyst type). The ranges and number of the corresponding input variables are summarized in Table 1.

2.3 Machine Learning Algorithms

ML classification models could be used to identify and classify material and group or map them in terms of their properties (descriptors), which is the first essential requirement prior to any ML-based predictions. We use the Scikit-learn package in the ML modules.[42]

The ML algorithms employed for classification of electrocatalyst and product type include logistic regression (LR), linear discriminant analysis (LDA), k-nearest neighbors (KNN) classifier, and random forest (RF) classifier. In addition, we tried to classify groups of products by putting all possible products into two or three different larger groups of products. In order to compare the predictability of different models for finding missing data, four ensembles of ML algorithms were evaluated. The regression algorithms include Bagging Regression (BR), Gradient Boosting Regression (GBR), Random Forest Regression (RFR), and Extra Trees Regression (ETR). BR is an ensemble method that fits regressors on random subsets of the original dataset and makes a final decision based on aggregated prediction. The bagging method increases the robustness of the original set of models by introducing randomness during the training process and then ensembling their predictions. GBR builds a model in a forward stage-wise style, which enables optimization on any differentiable loss functions. RF is a typical ensemble learning model that operates by building a set of decision trees and yielding average prediction of a separate tree. Random decision forests are superior to decision trees due to the ability to solve the over-fitting issue. Finally, extra trees implement a meta-estimator that fits several random decision trees on different sub-samples of the dataset and utilizes the mean of trees to boost the predictive performance and reduce the variance. ETR and RFR models have shown to be promising in the modeling of chemical systems. Each algorithm was trained on the training data for the CO_2 reduction reaction. The algorithms were then implemented to predict faradaic efficiency, applied potential (AP), and current density for the test dataset. We used the ML hyperparameter optimization module to tune hyperparameters automatically.

The accuracy score (%) (i.e., the ratio of correct predictions to the total number of predictions) is used as a performance metric for the evaluation of each classification algorithm. The performance of each ML algorithm for prediction was evaluated by using several statistical

indicators such as the mean squared errors (MSE), the root mean squared error (RMSE), and the coefficient of determination (R^2),

$$MSE(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2 \qquad (1)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2} \qquad (2)$$

$$R^2 = \frac{\sum_{i=0}^{n-1}(\hat{y}_i - \bar{y}_i)^2}{\sum_{i=0}^{n-1}(y_i - \bar{y}_i)^2} \qquad (3)$$

in which y_i and $\hat{y}_i$ are the true and predicted values, respectively, $\bar{y}_i$ is the mean of the true values, and n is the number of samples.

2.4 Modular Design

The complexity of the materials design-to-device integration calls upon a modular approach, in which various data management tasks and data analytics tools are built and tested in isolation as stand-alone-modules. The suitable modules are then called and integrated to the main platform depending upon application area, required analysis tools, and type of meta-data that the user needs to utilize for the analysis. In the following, we describe the adaptation of each module and their inter-dependencies for the analysis of electrocatalytic materials for CO2RR.

2.4.1 Classification and Materials Data Extraction

This module utilizes a classification algorithm that categorizes catalyst materials in the form of performance range (e.g., potential or current density) or selectivity or type of products. The reference values for high-level technical targets are based on a "performance matrix" that is provided as default for that field of application or as a user-entry table for the target values. These initial values can be seen as the first set of keywords for data mining and data discovery from the literature sources for given material application fields or sub-classes therein such as low-temperature catalysts, high-temperature catalyst. The extracted data is then mapped on

these key technical parameters and other crucial measurement conditions for each class of materials.

2.4.2 Materials Property Prediction

This module can predict a specific electrocatalytic property such as the faradaic efficiency as a function of input or exploratory variables using embedded ML models. The results of these ML prediction models can refine the usefulness and relevance of the user input variables. The module also helps fill missing data points related to performance indicators or target properties in the database as much as possible and thus enrich the master database. In this context, electrode type, current density, voltage, polarization resistance, conductivity, electrolyte type and composition, temperature, type of product, and (rarely) faradaic efficiency are among the key factors that may influence CO_2RR performance.

2.4.3 Recommendation System and Decision Models

The performance tuning algorithm is the first layer of the recommendation module that uses the complete dataset to find the best electrocatalyst material based on performance and stability metrics's target values. It displays the information using standard visualization tools, for example, using a radar chart. A radar chart is a typical visualization tool implied in benchmarking electrocatalyst material for the purpose of quality and performance improvement of a system of materials or an electrochemical device.[41] The use of radar charts makes two significant contributions: first, it provides a simple 2D visual representation of multiple performance indicators without the need of using dimensionality reduction on multivariate data, second, the surface area, formed by spikes (or axes), can be referred to as an electrocatalyst performance indicator.

The ML-powered recommendation module uses the power of regression modeling to predict values for the missing data as accurately as possible. **Supplementary Table 1** (Supplementary Materials) shows the sample data statistics used for training the regression models for predicting the missing data, specifically for applied potential, current density, and faradaic efficiency. Datapoints for four types of electrocatalyst material were selected, as there was not enough data for predicting other variables in the CO_2 experimental database.

2.4.4 Data Matching and Validation

The ultimate criterion for ML-based predictive capabilities is experimental validation, which demonstrates how computer algorithms lead to real discoveries. After predicting the best candidate electrocatalyst material, the prediction can be validated by direct comparison to experimental data for the same or almost the same set of conditions and catalyst materials specifications. [36]

In our predictive algorithm, CO_2 electrocatalyst materials are generally categorized into three main groups: metallic, non-metallic, and molecular catalysts. Each category of electrocatalyst materials exhibits distinct physicochemical and electrocatalytic properties. Therefore, it is possible that the performance of an electrocatalyst material is restricted and limited to the group of catalyst materials it belongs to. Here, we used ML classification models to classify different electrocatalyst materials into different groups based on their performance. The numerical data are normalized between 0 and 1, and we encoded the categorical data using "*OneHotEncoder*" from the Scikit-learn data preprocessing module. [42]

Most of the data in our Master database at low temperature is for Cu electrocatalyst, with the key properties of applied potential (AP), current density (CD), and faradaic efficiency (FE), type of electrolyte, and type of product. Material properties predictions thus focus on these attributes.

3. RESULTS and DISCUSSION

3.1 Materials Recommendation

Figure 2 shows the workflow of material recognition. In order to identify an electrocatalyst material for a given electrochemical process, it is expected that the performance metrics of the chosen electrocatalyst meet or exceed the target values set by the user. For this purpose, one needs to consider the key performance metrics, i.e., faradaic efficiency, current density, applied potential, selectivity, and production rate, while selecting the best electrocatalyst material.

In practice, keeping track of all variables and establishing correlations among optimization parameters in an electrochemical reaction path is a difficult task; once a set of properties is set to the optimum values, other properties of the catalyst can have values which below user

requirements. We attempt to address this challenge by introducing a penalty function for any value less than the desired value for a target application variable.

The recommendation process shown in Figure 2 initially takes input from the user-specified target values. The recommendation then selects the "best" electrocatalyst or recommends electrocatalysts materials primarily based on the targets for the set of performance metrics defined by the user. Global target values are provided as default if no user-entry target values are available. In order to minimize the optimization effort and for fast and better identification criteria, the user is provided with one of the following identification schemes: (i) find any electrocatalyst material for some desired value of a metric, with any chemical product. (ii) find any electrocatalyst material for target value metrics for a specific set of chemical products, (iii) find some desired performance metrics, within specific electrocatalyst material groups, with any product. (iv) find some desired properties within specific electrocatalyst material groups, for a certain set of chemical products. The user is given target values for selected metrics, electrocatalyst type, and chemical products, where a user is able to filter data based on products and electrocatalyst material or simply select all the possibilities. If the user provides target values for all metrics, the recommendation algorithm selects electrocatalyst material with properties equal (with less than 10% deviation) or better than the user target. If the user provides target values for a few properties and not all the properties, then the algorithm uses default global target values for those target properties that are not provided by the user.

Here, a rather simplified, yet straightforward method for selecting an electrocatalyst material is employed by using a radar chart to identify the material, which encloses the graph's maximum area. Although this heuristic method can be seen as practically useful, it may lead to a biased selection of electrocatalyst materials with few performance indicators at high values, while others remain at low values. It ignores the ranking and importance of different variables.

Our optimization algorithm employs a special scoring factor where it scores the positive value for properties that are higher or equal to the user target values and penalize properties that are less than the user target values. The value of the penalty function becomes more valuable for performance indicators that significantly less than the actual target values. This sub-routine recommends catalyst materials that exhibit high values in one or multiple attributes from performance matrix table.

The scoring factor is defined by,

$$Score_i = \sum_{m=1}^{m=k} difference_{im} \qquad (4)$$

$$i: is\ the\ number\ of\ row$$

where k is the number of target properties (P). If $P_{ij} \geq P_{User\ Target j}$

$$difference_{ij} = P_{ij} - P_{User\ Target j}$$

If $P_{ij} < P_{User\ Target j}$

$$difference_{ij} = 5^{(P_{ij} - P_{User\ Target j})} - 1$$

Where P_{ij} is the default target value of property j for the row number i and $P_{User\ Target j}$ is user defined target value for property j.

The constraint for the penalty function is set at 5, representing the maximum error tolerated. Once the scoring factors for each row in the database are calculated, the algorithm recommends electrocatalyst materials with high score values, as illustrated in **Supplementary Figure 1**.

3.2 Low-temperature electrocatalyst materials

Figure 3 shows the visualization of data, which is distributed among applied potential, current density, and faradaic efficiency for different types of electrocatalysts at low- or high-temperature. The diagonal graphs represent the density plot of each respective feature, providing useful information by giving a density of plots in the form of bar charts. Among the possible choices of electrocatalysts at low temperatures, mainly four types of Cu-based electrocatalysts are used for the classification task. The dataset is divided into training and test datasets. The dataset consists of 228 of different Cu electrocatalyst materials, in which training and test datasets consist of 183 and 45 data points, respectively. Each data point consists of a set of properties for a given material. The same materials may appear in different data points with different operating conditions. The materials space is then scanned using a set of identified descriptors, such as selectivity for a given product or performance indicators against a reference target range. The latter is performed using machine learning techniques. Model performance for classification of the type of electrocatalyst and type of products was evaluated through the calculation of an accuracy score.

As illustrated in Table 2, the key indicators (AP, CD, FE, Product selectivity) have high cross-validation scores, which can vary according to the ML algorithms. The LR and LDA classifiers are found to return the highest accuracy score of 81%, determining the type of electrocatalyst. QDA classifier has an accuracy score of 32%, which is remarkably lower than that for other classifiers.

As shown in Table 3, the indicators of AP, CD, FE, and type of electrocatalyst yield a higher accuracy for classification of a group of two products [CH_4, C_2H_5OH] in comparison with two other groups, each consists of three different products. RF and LDA classifiers return value of 1 and 0.93, respectively, for accuracy score of all test cases. In general, RF classifier has the best performance among other algorithms for the classification of the type of products regardless of the number of products.

LR, LDA, QDA, and GNB algorithms were unable to distinguish and single out one group of products, including those with three different products. Additionally, GNB returns an accuracy score of 26%, the lowest of all six algorithms. It is obvious that the better performance of ML algorithms can be achieved for the group of two products than the groups with three different products. The latter can be understood from the comparison of the values of accuracy score for classification of the type of electrocatalyst or products reported in Tables 2 and 3. One would need more indicators such as the reaction conditions (pH, mass loading of catalyst, production rate, and concentration) for each reaction in order to have a better performance with the classification scheme.

Table 4 lists the performance of predictive analytics using MSE for various experimental numerical values, i.e., applied potential (AP), current density (CD), and faradaic efficiency (FE). ETR is seen to have a better predictive capability with a minimum error, which is considered more accurate than other algorithms. In order to quantitively obtain a prediction model for FE, AP and CD, BR, GBR, ETR, and RFR algorithms were employed. Models were based on the training data (80% of the full dataset), where 20% of that is used to evaluate the test data.

The scatter plot of the outputs versus the actual values for the training, testing, and overall data sets using RFR and ETR algorithms are presented in Figure 4. The coefficient of determination (R^2) indicates a strong correlation between outputs for CD and AP and actual values. The AP,

CD, and FE results clearly show excellent agreement between the actual values and RFR, GBR, and ETR predictions, with $R^2 > 0.90$ and MSE < 0.008 for all of the ensemble modeling cases. The R^2 and MSE of test data for faradaic efficiency with ETR and RFR have better performance than that for other regressors.

Success with ML agorithms depends on the number of descriptors and their correlations, as well as available large training data. The true benefit of structure-property relationships revealed through ML models lies in the multi-variant correlations and their interpretation in terms of the fundamental materials properties.

The missing values in the primary database can nonetheless be filled with values extrapolated from ML by building a model that relates the known indicators of materials to target properties. Our ML model has successfully predicted different properties like FE or CD, or classification of the type of electrocatalyst, or major products related to specific tyoe of catalyst. The latter process has been carried out iteratively. After filling missing values, the database is ready to screen the electrocatalyst performance through means of analytical and visualization tools.

Utilizing all available and supplemented databases, rapid screening of electrocatalyst material was carried out, while the user was able to specify target values for various properties. The optimization algorithm proposed in this work uses a scoring factor based on a rank-ordering approach. The best electrocatalyst material for given chemical products was then estimated for a class of materials or products. Figure 5 shows the radar charts of the best electrocatalyst material based on the target attributes selected by users or directly from a global target, which is set as default. The chart indicates that Pt should be the catalyst of choice when no specific fuel products are considered.

3.3 High-temperature Electrocatalyst Materials

Despite recent advances and development of electrolytic processes for CO_2 conversion at high temperature (> 800 ° C), the overall efficiency and performance of the system remain far from understood for commercialization and practical usage.[43] Among the technological shortcomings are low conversion efficiency and high degradation rates of materials and components, including membrane and electrocatalysts. The latter is mainly due to the fact that the high catalytic conversion will inherently result in low electrochemical stability of catalyst materials at higher temperatures. The fundamental understanding of the elementary kinetic processes involved in CO_2 electrochemical conversion at high temperatures is a subject of ongoing research.[34] Notably, the cost-effectiveness of the catalytic process at high temperatures primarily depends upon trade-offs between the system efficiency and production cost of the fuel, while the operating condition of the solid oxide electrolyzer cells (SOECs) remains very narrow due to high heat requirements and sensitivity to temperature fluctuations.[35] CO is the major carbonated product as all other competing chemical reaction products are desorbed from the surface to produce CO at high temperatures. Therefore, additional down-stream processes need to be considered in order to achieve other products such as methanol. For co-electrolysis of CO_2 and H_2O, SOEC provides high flexibility in the carbon to hydrogen ratio (C/H) and, thus, state-of-the-art technologies such as Fischer Tropsch (FT) synthesis can be utilized at downstream for achieving high product flexibility. [45,46]

Here, we present preliminary results and discussions for data-driven analysis of selected electrocatalyst systems in SOECs that addresses a few of the above technological challenges. In high-temperature electrolysis of CO_2, the co-electrolysis process in the presence of steam is taking place at temperatures >600 °C. High-temperature CO_2 electrochemical conversion using SOEC generally has a better selectivity compared to that at low temperatures. Correlations among AP, CD and FE at low or high temperature are not known yet and require extensive data analytics.

The state-of-the-art high-temperature electrocatalyst materials in SOECs contain Ni-YSZ. A key factor for the stability and activity of these materials at high-temperatures is Ni% in the range of 40–60%. This range is required to fulfill the catalytic reforming and satisfies the thermal expansion coefficient match between the catalyst layer and YSZ electrolyte. [43] Similar

to solid oxide fuel cell (SOFC) electrodes, electrocatalytic reactions in SOECs take place at triple phase boundaries (TPB) where the Ni phase provides electrons, and YSZ particles offer the required oxygen ion vacancies for the reduction of adsorbed CO_2 and the removal of the separated oxygen ions, respectively.

Recent progress and advancement for the high temperate electrochemical reduction of carbon dioxide suggest that the electrochemical reduction of CO_2 in solid oxide electrolysis cells takes place at high current densities. Degradation rates are higher in electrolysis mode compared to those in fuel cell mode based on new or enhanced issues such as metal particle migration and/or oxidation, carbon deposition, grain coarsening, and impurities contamination. This adds complexities to the choice of electrocatalyst materials and, thus, significant fundamental research activities. In particular, electrochemical reduction of CO_2 in the temperature range of 573–873 K is worth exploring in order to match the temperature levels of electrolysis with required downstream FT-processes; however, there are no proper material systems for the electrodes and electrolyte under that temperature regime at the current stage.

Here we consider a few conventional classes of electrode materials and explore the impact of ratios and Ni, or Ti on overall catalytic activity via extensive data analytics.

Figure 3b provides scatter plots and distributed values for applied potential, current density, and faradaic efficiency for Ti and Ni-YSZ catalyst systems. Ti-based electrocatalyst exhibits different dependencies for applied potential and faradaic efficiency compared to that for the Ni-YSZ system, while both catalyst materials are relatively similar correlations in view of current densities. Overall, the Ti-based catalytic system shows high correlations among FE and CD, in particular in the range of data obtained at higher applied potentials ranges (> 2 V). Figure 3a and 3b clearly reveal differences in the correlations among key attributes such as FE and AP among catalysts at low and high temperatures. The correlations are more pronounced among FE and applied potential for high-temperature electrocatalyst, whereas CD and AP are the main indicators at low temperatures. Among all electrocatalyst materials studied at high temperature, Ni-YSZ shows the highest correlation between FE and AP, although the correlation factors can vary depending upon Ni ratios and type of electrolytes or products, as illustrated in binary correlations in Figure 6.

The dataset for high-temperature catalysts consists of 180 test data points distributed among five different catalysts types. This amount of data is insufficient for accurate prediction of missing properties in the data set, and thus further predictions using ML techniques and identification thereof are not feasible based on the existing size of the dataset. Moreover, the atomic ratios of the composite electrocatalysts are not taken into consideration in these databases. The current results, however, will be expanded in the future to further insights for the correlation of key attributes at high-temperatures using larger and more diverse training and test data sets.

3.4 Recommendation and Decision System

Here, we only focused on high-level correlations among selected indicators. **Supplementary Table 2** (Supplementary Materials) provides the complete test data and other operational conditions that are assumed for each data point. The type of electrolyte is another important factor to be considered as it influences the extend of correlations among FE and AP for various high-temperature electrolysis technologies and the respective electrocatalysts. In particular, future work can include the analysis for the following use cases and comparison based on phase ratios and catalyst types such as Ag, Ni|YSZ or Ag|YSZ and for at least one cell configuration such as Ag/GDC|YSZ|YSZ/LSM|LSM (La 0.8 Sr 0.2 Cr 0.5 Mn 0.5 O 3 $-\delta$(LSCM)). Further analysis is still ongoing to improve the test and training databases for high-temperature catalyst and provide a robust recommendation framework for this system. Here, the analysis is primarily built upon existing and extracted historical data. There is an emerging need for employing sophisticated decision algorithms and recommendation systems to “close-the-loop”. Such algorithms are emerged from predictive models of key materials properties under different experimental conditions or modeling assumptions. They also identify weighting factors that govern specifications and limitations imposed at the components and device-level integration of new materials. Such algorithms are trained over time as more historical data and use cases become available.

4. Conclusions

The discovery and optimization of electrocatalyst materials are driven in large part by collecting and analyzing various experimental data. The ML-assisted development of real electrocatalysts is still an emerging field despite its success in molecular and material science; it cannot yet lead directly to novel electrocatalyst design.

In this article, we proposed a recommendation framework for the benchmarking of existing electrocatalyst materials. A multi-attribute decision process was adopted, which was mapped on radar charts, from which the analysis of best-performed electrocatalyst is carried out based on user-entry or global technological targets. This recommendation framework provides the choice of dimensions, indicators, and appropriate correlations for benchmarking purposes and for assessing the electrocatalyst materials screenings process, purely based on historical data. With the availability of reliable process and materials cost data, the latter can lead to comprehensive techno-economic insights into what performance levels are required for commercially viable electrocatalytic reactions within the clean energy sector.

We used ML to supplement missing values in CO_2RR databases prior to deploying ML algorithms to identify the best catalytic system with the highest overall performance. The ML module is primarily built for the classification and prediction of electrocatalyst materials. Different models for classification of the type of electrocatalyst materials and chemical products are used with reasonable accuracy within the limit of available test and training data. Among different regression algorithms, the Random Forest model had a better capability for the prediction of electrochemical indicators. The proposed recommendation system provides interactive visual analysis of different indicators for the exploration of uploaded electrocatalyst data. High-level correlation analytics was also provided for catalyst materials at high temperatures, and the intensity of correlations are compared to that for catalyst materials at low temperature.

Finally, rapid screening and benchmarking studies of electrocatalyst material via data-driven visualization can significantly reduce the discovery time for the best catalyst materials and to understand or compare vital performance trends and correlations for given classes of materials from initial discovery, to component or device integration and for full-scale component or device production. The major limitations of the framework presented here are the lack of available datapoints, un-clarity or lack of consistency around key numerical or categorical attributes, and missing values for the attributes that are collected from the literature. The framework, however, can be applied to other sustainable electrochemical processes such as electrochemical NH_3 synthesis through N_2 and H_2O electrolysis.

The interactive visualization tools, assist researchers in discovering trends and patterns hidden with the electrocatalyst material based on historical experimental and modeling data. Further ML and analytics functionalities are currently under development, which will offer higher accuracy and better inter-operability of the recommendation

framework for idea-creation and screening state-of-the-art electrocatalyst materials for various applications.

Conflict of Interest

There are no conflicts to declare.

Acknowledgments

This work was supported by the German-NRC collaboration project. KM and QW would like to thank NRC international office and NRC's Materials for Fuel Challenge program for their financial support. The authors also greatly acknowledge researchers in IEK-13 and IEK-1 at the Forschungszentrum Jülich for valuable insights and contributions to this project. Contribution from Gagandeep Singh Bajwa at NRC-EME for the extraction, cleaning, and analysis of CO_2RR databases is greatly acknowledged. ME acknowledges the support from the Forschungszentrum Jülich GmbH.

Supplementary Materials

The actual data is collected from experimental CO_2RR measurements from a series of journal articles. The details of data retrieval from images, tables, and texts are described in ref [36]. Supplementary materials include “definistion of key attributes”, Supplementary Figure 1 and Supplmentary Table 1 that contains raw data in excel format. The underlying datasets related to this manuscript being available publicly from Github (*VMILabs/Data-Public*).

References

1. L. Al-Ghussain, *Environmental Progress & Sustainable Energy*, 2019, **38**, 13-21.
2. J. Rogelj, Michiel Schaeffer, and Bill Hare, *Berlin, Germany: Climate Analytics (http://climateanalytics.org/publications/2015/timetables-for-zero-emissions-and-2015-emissions-reductions)*, 2015.
3. J. Qiao, Y. Liu, F. Hong and J. Zhang, *Chemical Society Reviews*, 2014, **43**, 631-675.
4. Q. Lu and F. Jiao, *Nano Energy*, 2016, **29**, 439-456.
5. D. D. Zhu, J. L. Liu and S. Z. Qiao, *Advanced Materials*, 2016, **28**, 3423-3452.
6. X. Liu, J. Xiao, H. Peng, X. Hong, K. Chan and J. K. Nørskov, *Nature communications*, 2017, **8**, 1-7.
7. Y. Wang, J. Liu, Y. Wang, A. M. Al-Enizi and G. Zheng, *Small*, 2017, **13**, 1701809.
8. R. Lin, J. Guo, X. Li, P. Patel and A. Seifitokaldani, *Catalysts*, 2020, **10**, 473.
9. M. Mandal, *ChemElectroChem*, 2020.
10. M. G. Kibria, J. P. Edwards, C. M. Gabardo, C. T. Dinh, A. Seifitokaldani, D. Sinton and E. H. Sargent, *Advanced Materials*, 2019, **31**, 1807166.
11. A. Goyal, G. Marcandalli, V. A. Mints and M. T. Koper, *Journal of the American Chemical Society*, 2020, **142**, 4154-4161.
12. Y. Liu, K. Y. Leung, S. E. Michaud, T. L. Soucy and C. C. McCrory, *Comments on Inorganic Chemistry*, 2019, **39**, 242-269.
13. K. Elouarzaki, V. Kannan, V. Jose, H. S. Sabharwal and J. M. Lee, *Advanced Energy Materials*, 2019, **9**, 1900090.
14. H. Ju, G. Kaur, A. P. Kulkarni and S. Giddey, *Journal of CO2 Utilization*, 2019, **32**, 178-186.
15. B. Cao, L. A. Adutwum, A. O. Oliynyk, E. J. Luber, B. C. Olsen, A. Mar and J. M. Buriak, *ACS nano*, 2018, **12**, 7434-7444.
16. P. De Luna, J. Wei, Y. Bengio, A. Aspuru-Guzik and E. Sargent, *Journal*, 2017.
17. A. Varnek, N. Kireeva, I. V. Tetko, I. I. Baskin and V. P. Solov'ev, *Journal of chemical information and modeling*, 2007, **47**, 1111-1122.
18. B. R. Goldsmith, J. Esterhuizen, J. X. Liu, C. J. Bartel and C. Sutton, 2018.
19. G. R. Schleder, A. C. Padilha, C. M. Acosta, M. Costa and A. Fazzio, *Journal of Physics: Materials*, 2019, **2**, 032001.
20. P. S. Lamoureux, K. T. Winther, J. A. G. Torres, V. Streibel, M. Zhao, M. Bajdich, F. Abild-Pedersen and T. Bligaard, *ChemCatChem*, 2019.
21. A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer and C. Hargus, *Journal of Physics: Condensed Matter*, 2017, **29**, 273002.
22. K. Mathew, J. H. Montoya, A. Faghaninia, S. Dwarakanath, M. Aykol, H. Tang, I.-h. Chu, T. Smidt, B. Bocklund and M. Horton, *Computational Materials Science*, 2017, **139**, 140-152.
23. A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner and G. Ceder, *Apl Materials*, 2013, **1**, 011002.
24. C. Draxl and M. Scheffler, *Journal of Physics: Materials*, 2019, **2**, 036001.
25. C. Citrine Informatics, available at https://citrination.com.).
26. J. S. Hummelshøj, F. Abild-Pedersen, F. Studt, T. Bligaard and J. K. Nørskov, *Angewandte Chemie International Edition*, 2012, **51**, 272-274.
27. G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari and B. Kozinsky, *Computational Materials Science*, 2016, **111**, 218-230.
28. B. Meyer, B. Sawatlon, S. Heinen, O. A. von Lilienfeld and C. Corminboeuf, *Chemical science*, 2018, **9**, 7069-7077.
29. A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow and S. E. Denmark, *Science*, 2019, **363**.
30. A. Smith, A. Keane, J. A. Dumesic, G. W. Huber and V. M. Zavala, *Applied Catalysis B: Environmental*, 2020, **263**, 118257.
31. J. R. Kitchin, *Nature Catalysis*, 2018, **1**, 230-232.
32. P. Schlexer Lamoureux, K. T. Winther, J. A. Garrido Torres, V. Streibel, M. Zhao, M. Bajdich, F. Abild-Pedersen and T. Bligaard, *ChemCatChem*, 2019, **11**, 3581-3601.

33. S. Gusarov, S. R. Stoyanov and S. Siahrostami, *The Journal of Physical Chemistry C*, 2020, **124**, 10079-10084.
34. K. Tran and Z. W. Ulissi, *Nature Catalysis*, 2018, **1**, 696-703.
35. X. Ma, Z. Li, L. E. Achenie and H. Xin, *The journal of physical chemistry letters*, 2015, **6**, 3528-3533.
36. A. Malek, M. J. Eslamibidgoli, M. Mokhtari, Q. Wang, M. H. Eikerling and K. Malek, *Chemphyschem : a European journal of chemical physics and physical chemistry*, 2019, **20**, 2946-2955.
37. T.-C. Chou, C.-C. Chang, H.-L. Yu, W.-Y. Yu, C.-L. Dong, J.-J. Velasco-Vélez, C.-H. Chuang, L.-C. Chen, J.-F. Lee and J.-M. Chen, *Journal of the American Chemical Society*, 2020, **142**, 2857-2867.
38. A. J. Garza, A. T. Bell and M. Head-Gordon, *Acs Catalysis*, 2018, **8**, 1490-1499.
39. K.-Y. Chan and C.-Y. V. Li, *Electrochemically enabled sustainability: devices, materials and mechanisms for energy conversion*, CRC Press, 2014.
40. S. Nitopi, E. Bertheussen, S. B. Scott, X. Liu, A. K. Engstfeld, S. Horch, B. Seger, I. E. Stephens, K. Chan and C. Hahn, *Chemical reviews*, 2019, **119**, 7610-7672.
41. R. Basu, *Implementing quality: a practical guide to tools and techniques: enabling the power of operational excellence*, Cengage Learning EMEA, 2004.
42. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss and V. Dubourg, *the Journal of machine Learning research*, 2011, **12**, 2825-2830.
43. R. J. Gorte, S. Park, J. M. Vohs and C. Wang, *Advanced Materials*, 2000, **12**, 1465-1469.
44. Martha M. Flores-Leonar, , Luis M. Mejía-Mendoza, Andrés Aguilar-Granda, Benjamin Sanchez-Lengeling, Hermann Tribukait, Carlos Amador-Bedolla, and Alán Aspuru-Guzik.*Current Opinion in Green and Sustainable Chemistry* 25 (2020): 100370.
45. X. Zhang Xiaomin Zhang, Yuefeng Song, Guoxiong Wang, Xinhe Bao Journal of Energy Chemistry 26 (2017) 839-853
46. Y. Zheng Yun Zheng, Jianchen Wang, Bo Yu, Wenqiang Zhang, Jing Chen, Jinli Qiao, and Jiujun Zhang *Chem. Soc. Rev.,* 2017, **46**, 1427--1463

Figure 1. The workflow of the cognitive material identification system

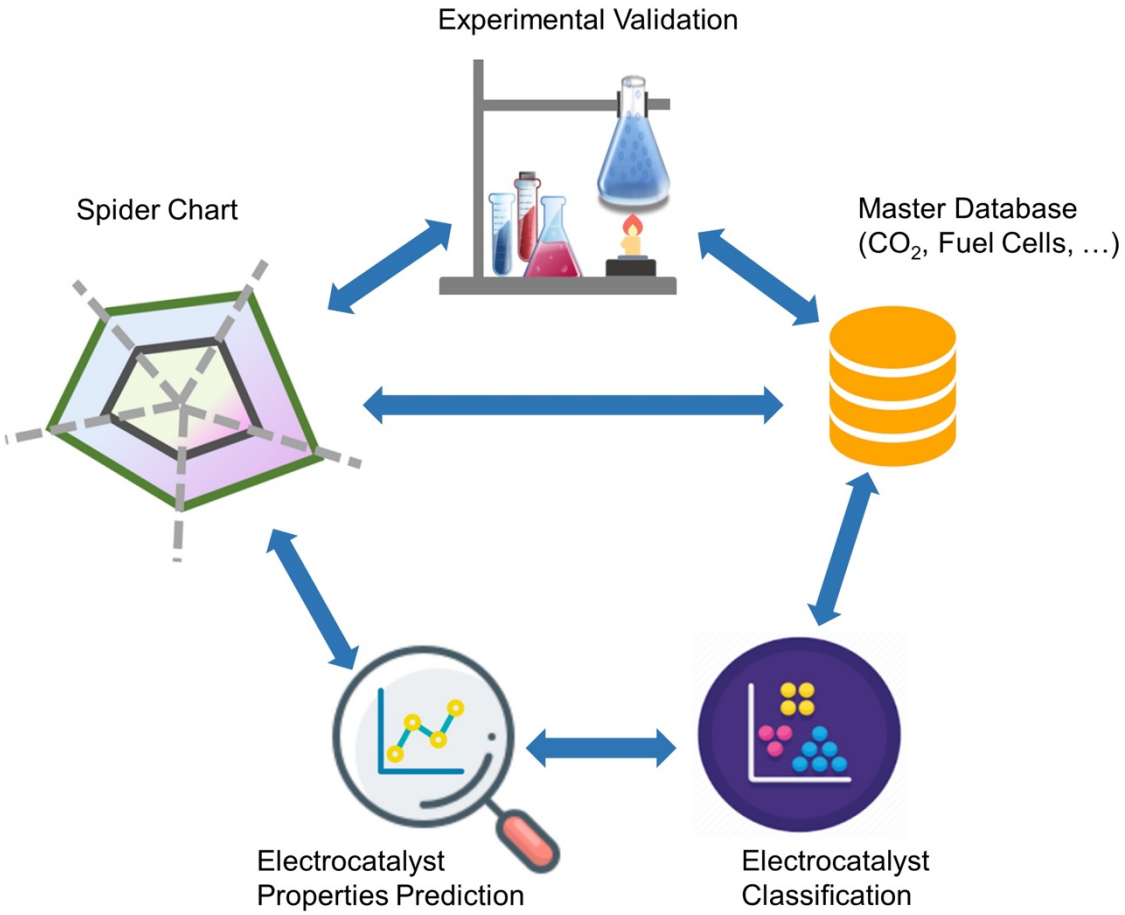


Figure 2. Flowchart of material identification framework

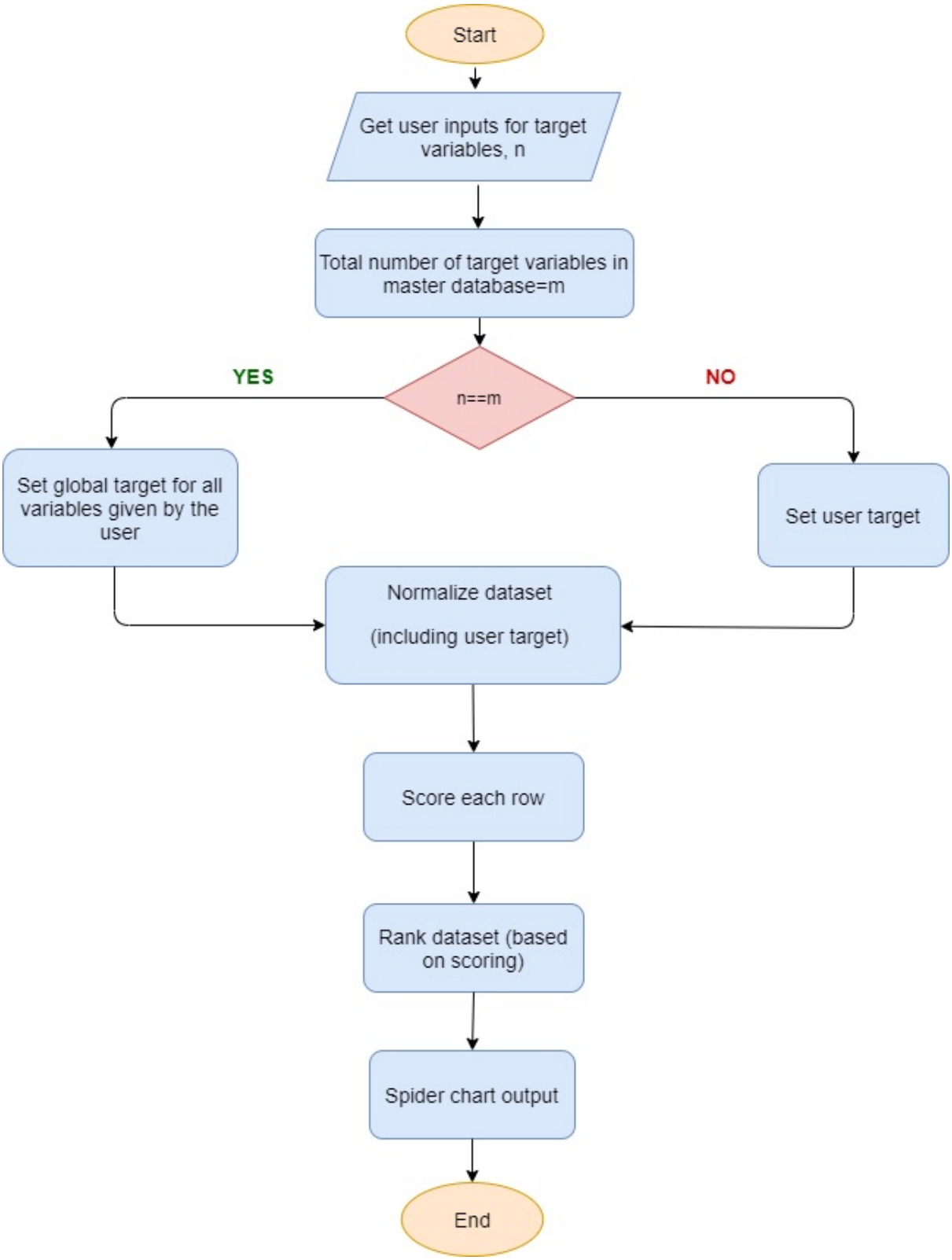
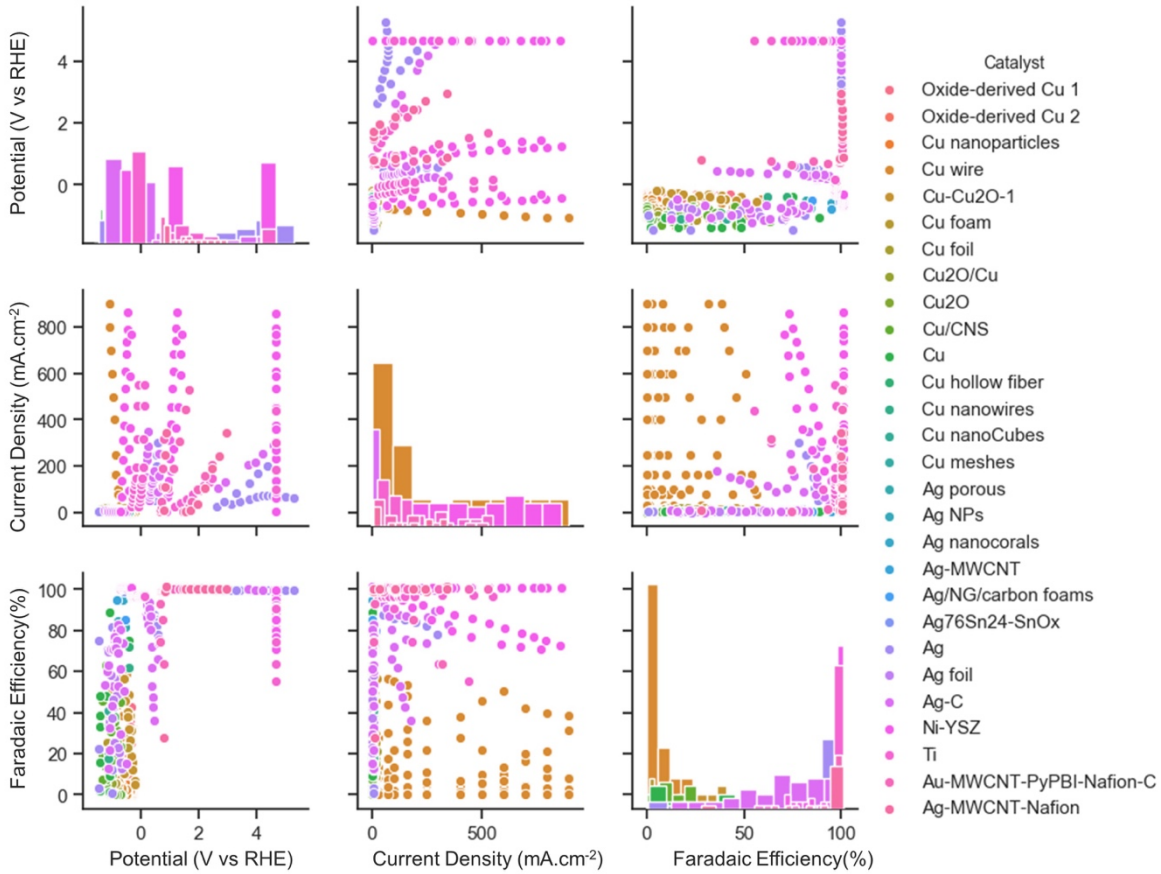
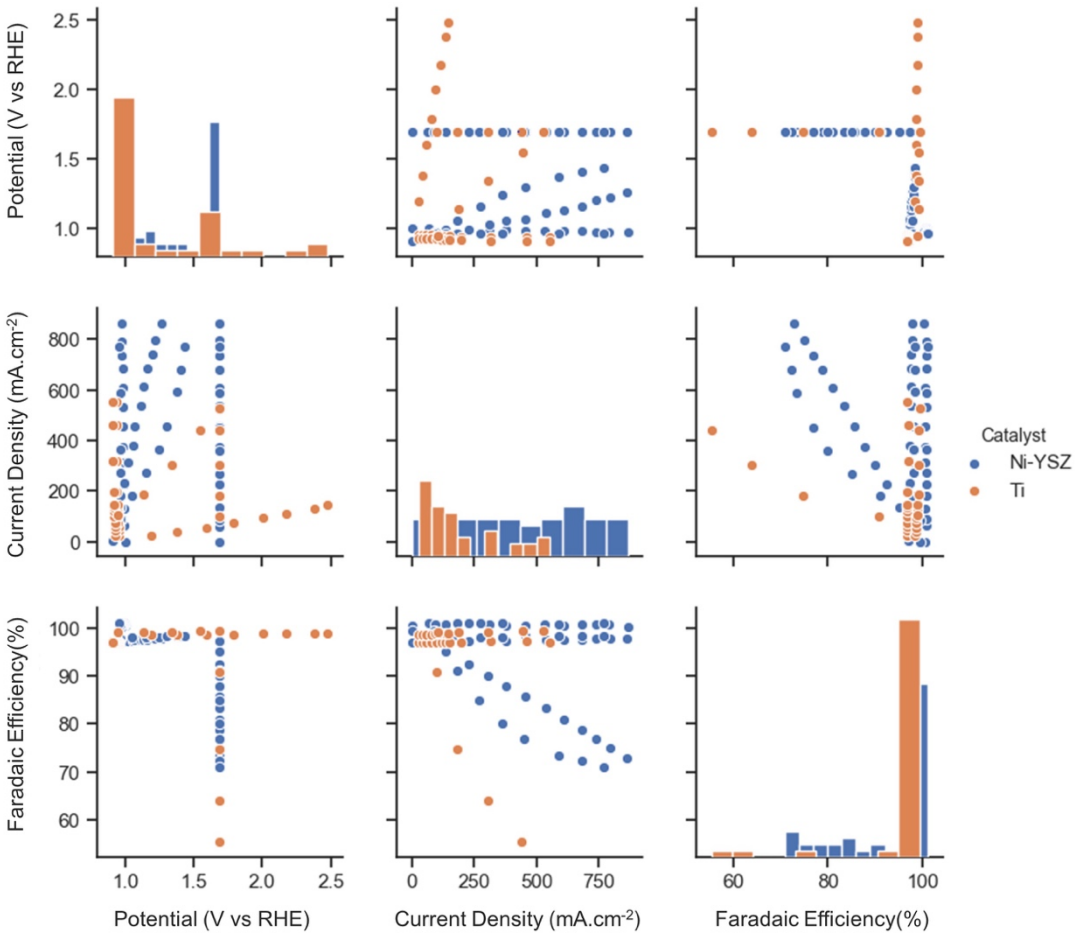


Figure 3. Scatter plot matrix showing the data distribution for a) both High-T and low-T b) High-T, c) Low-T of CO_2RR according to 3 performance metrics. The elements in the diagonal (upper left to lower right) represent the respected range of data points for each catalyst type.

(a)



(b)


Potential (V vs RHE)
Current Density (mA.cm-2)
Faradaic Efficiency(%)
Catalyst
Ni-YSZ
Ti

(c)

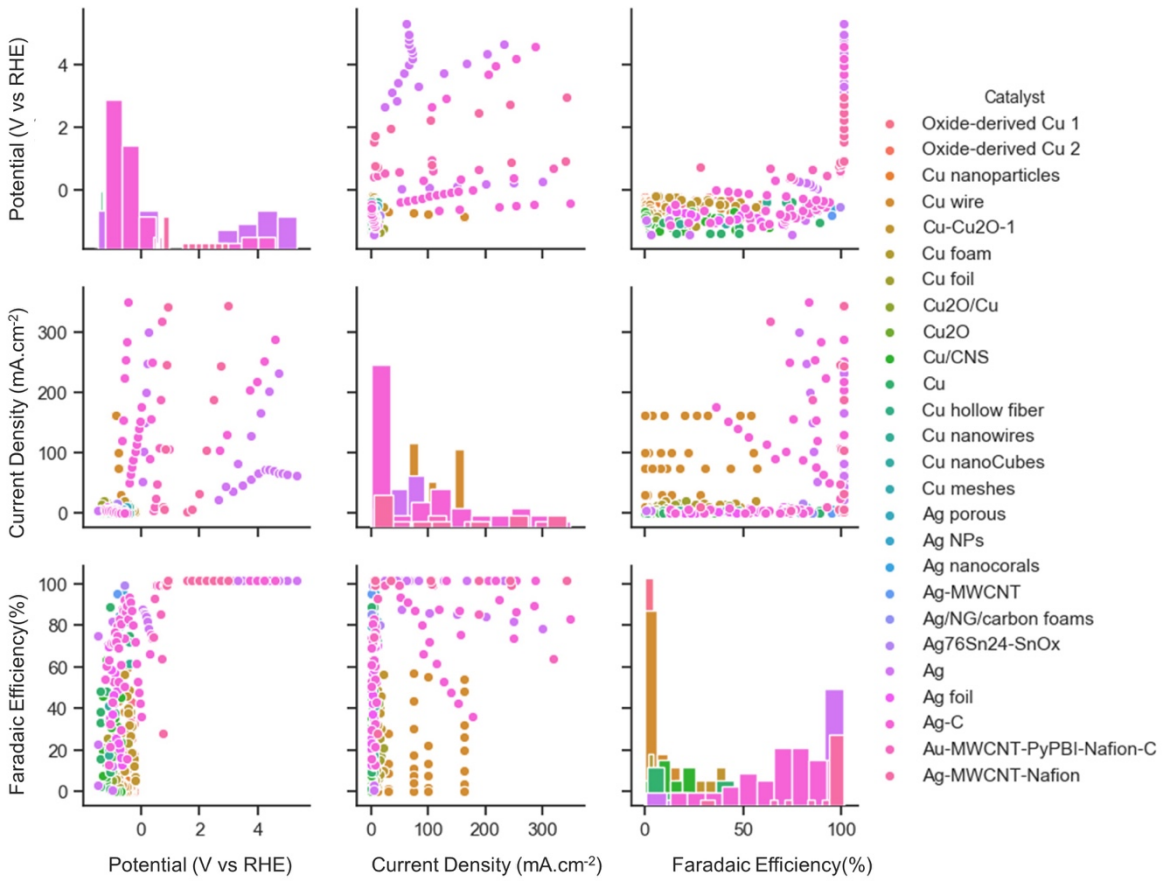

Potential (V vs RHE)
Current Density (mA.cm-2)
Faradaic Efficiency(%)
Potential (V vs RHE)
Current Density (mA.cm-2)
Faradaic Efficiency(%)
Catalyst
Oxide-derived Cu 1
Oxide-derived Cu 2
Cu nanoparticles
Cu wire
Cu-Cu2O-1
Cu foam
Cu foil
Cu2O/Cu
Cu2O
Cu/CNS
Cu
Cu hollow fiber
Cu nanowires
Cu nanoCubes
Cu meshes
Ag porous
Ag NPs
Ag nanocorals
Ag-MWCNT
Ag/NG/carbon foams
Ag76Sn24-SnOx
Ag
Ag foil
Ag-C
Au-MWCNT-PyPBI-Nafion-C
Ag-MWCNT-Nafion

Figure 4. The actual Faradaic efficiency, applied potential, and Current density values compared with the predicted values using Random Forest regression and Extra Tree regression models. The coefficient of determination (R^2) and mean squared error (MSE) are computed to estimate the prediction errors. The test and training points are shown in blue and red, respectively. The perfect correlation line is included for reference as a green line.

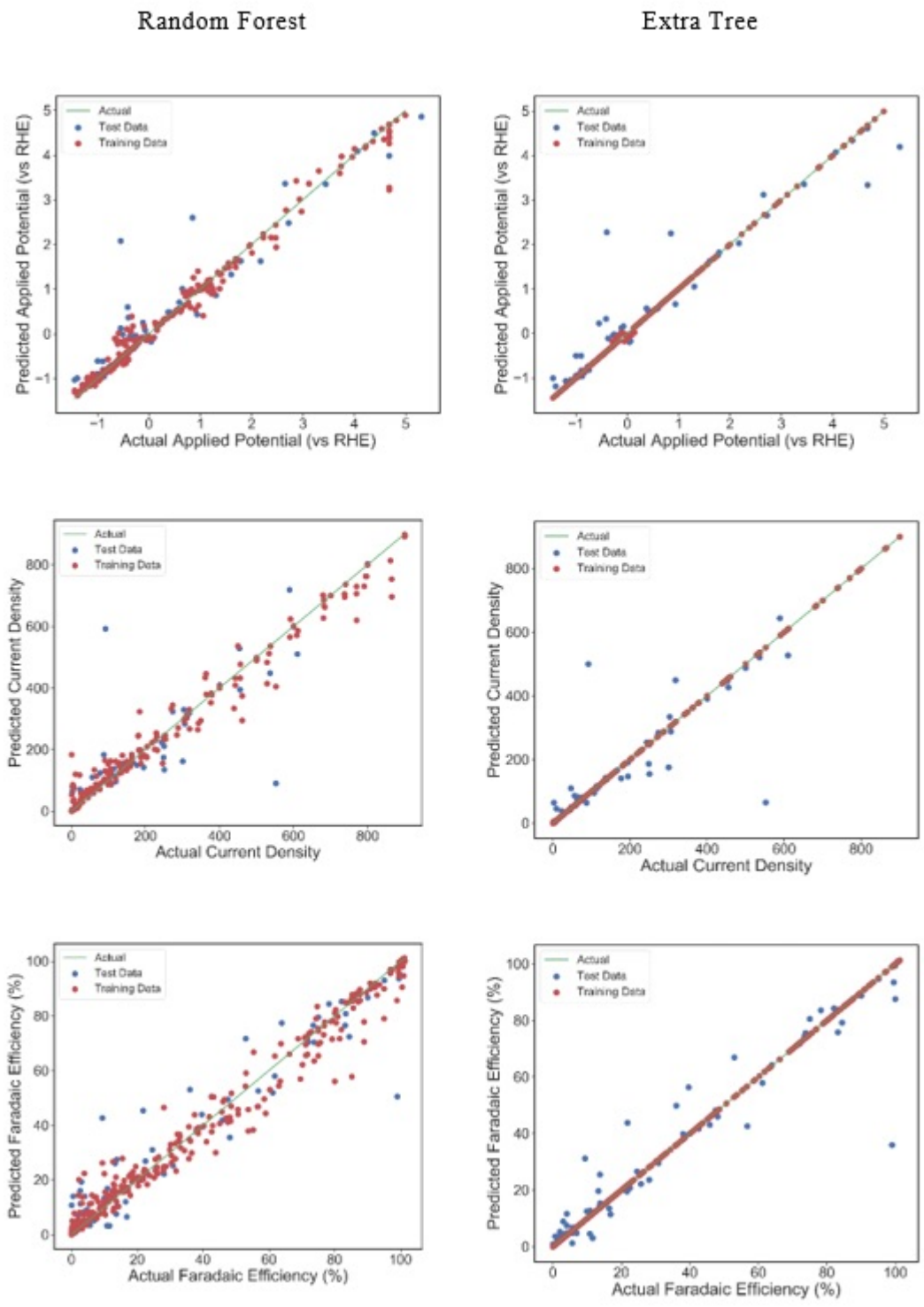


Figure 5. Screenshot of radar chart for CO_2 reduction to fuels of Pt-based on different classification of electrocatalysts and selected target values by default global (blue) or user-entry (brown) targets.

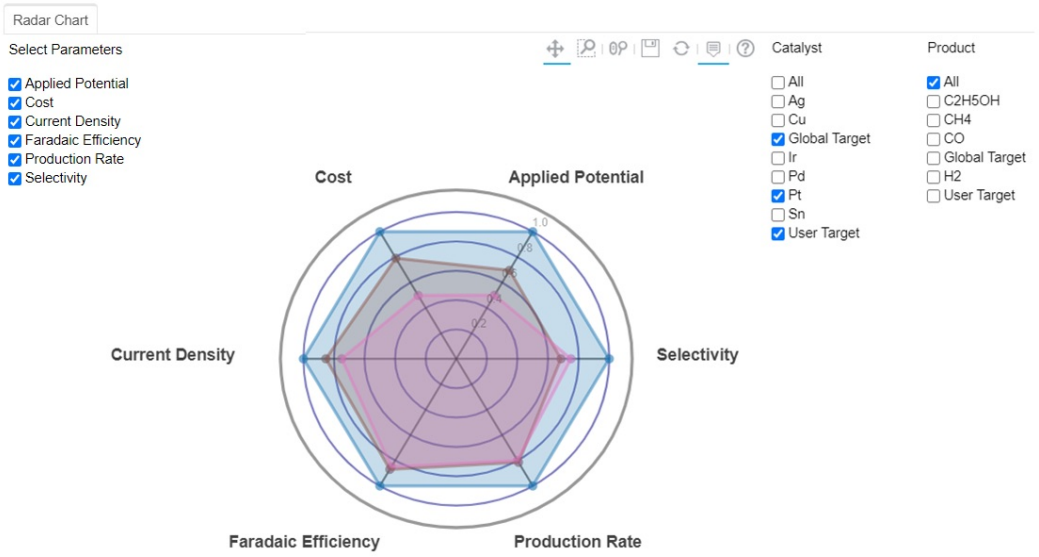
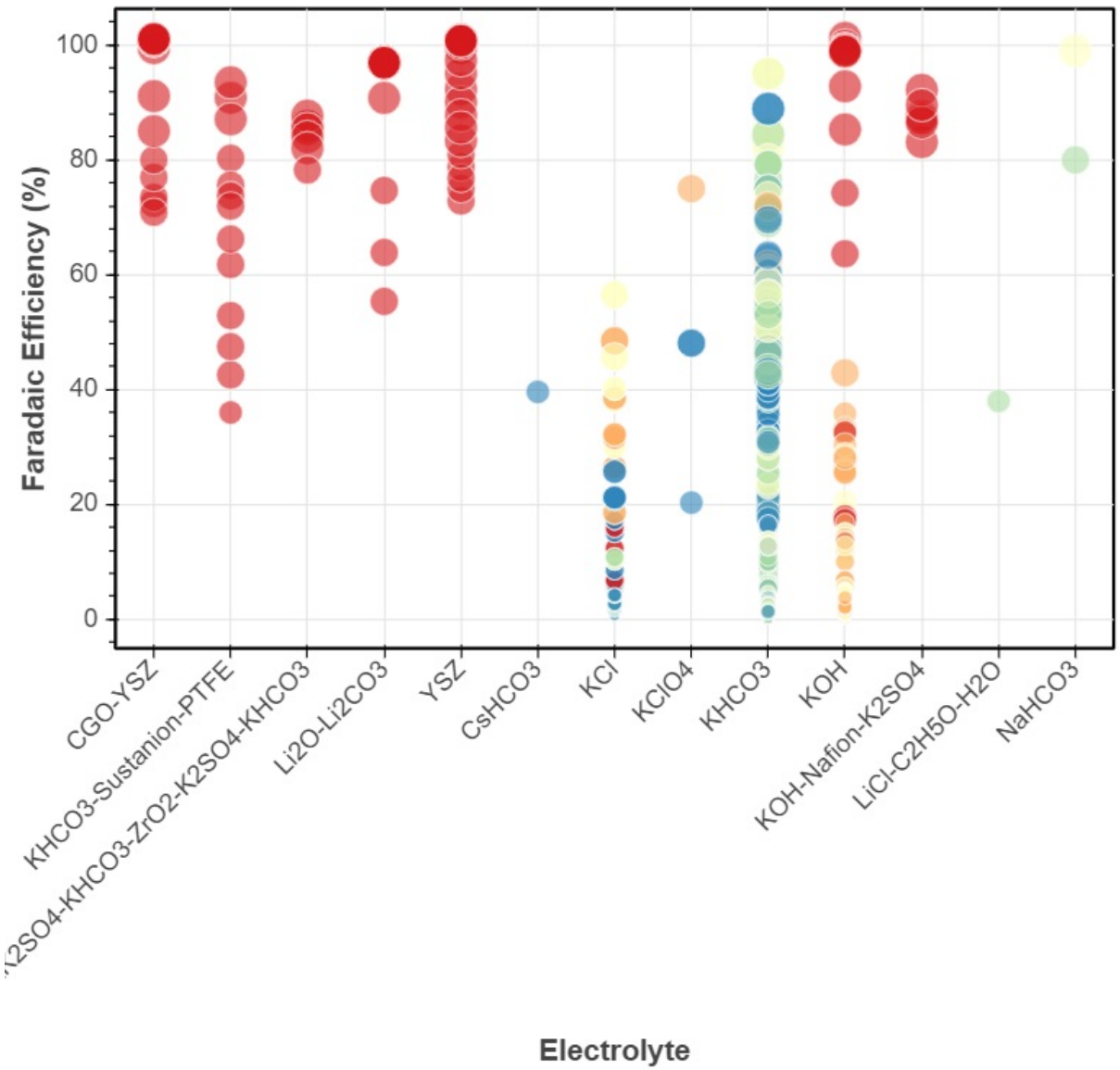
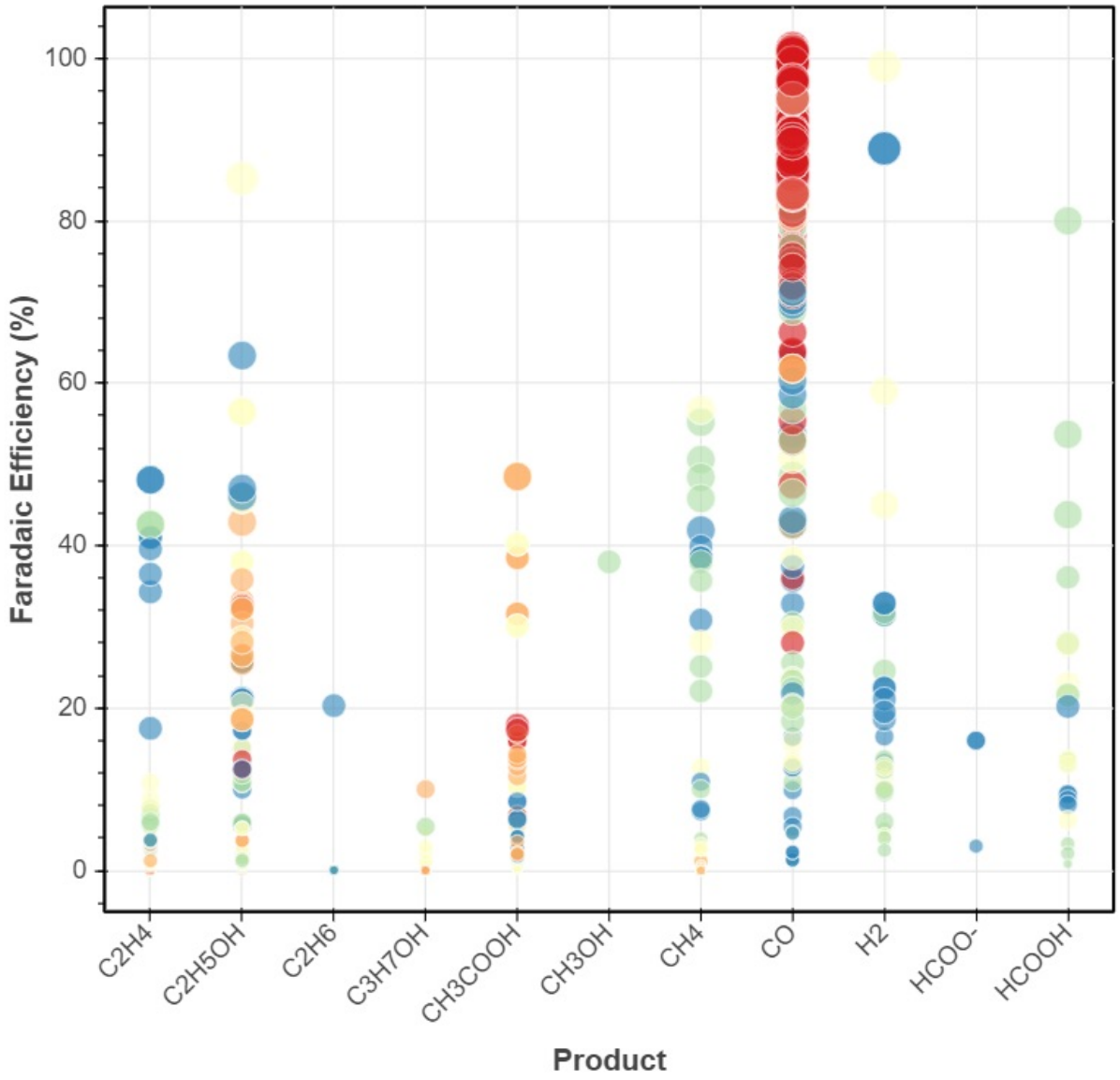


Figure 6. Binary correlations among key attributes (a) FE-electrolyte, (b) FE-product, (c) FE-current density for sample extracted datasets of electrocatalyst materials at low and high temperature.

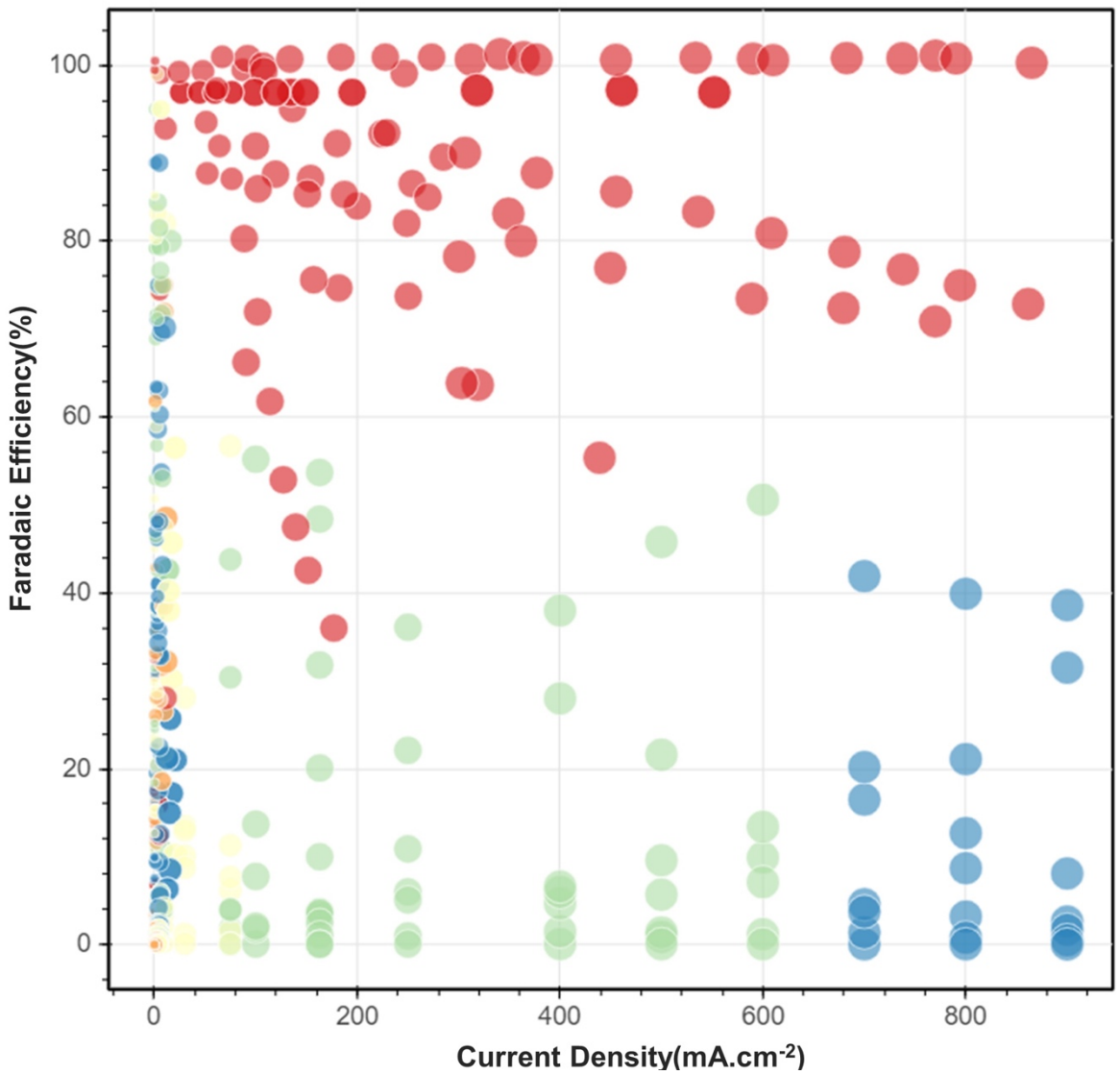
(a)



(b)


Faradaic Efficiency (%)
100
80
60
40
20
0
C2H4
C2H5OH
C2H6
C3H7OH
CH3COOH
CH3OH
CH4
CO
H2
HCOO-
HCOOH
Product

(c)


Faradaic Efficiency(%)
100
80
60
40
20
0
0
200
400
600
800
Current Density(mA.cm-2)

878 **Table 1.** Key performance indicators and their types or range of values as being set in the
879 data extraction process.
880

Descriptors	Range or types	Units
Catalyst	Cu, Ag, Ni, Ti	
Applied Potential	-1.45 to 5.3	V
Current Density	0.00058-856	mA/cm^2
Faradaic Efficiency	0 to100	-
Type of Electrolyte	KOH, KCl	
	$KHCO_3$,	
	$CsHCO_3$,	
	YSZ, Li_2O-Li_2CO_3	
Major products	CO, H_2,	
	CH_3COOH	
	C_2H_5OH,	
	C_3H_7OH	
Temperature	25 to 900	˚C

881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897

Table 2. The results of cross-validation with six different classification algorithms against low-temperature catalyst types in four classes (Cu wire, Oxide-derived Cu 1, Oxide-derived Cu 2, Cu nanoparticles)

Catalysts: Cu wire, Oxide-derived Cu 1, Oxide-derived Cu 2, Cu nanoparticles

ML Algorithms	Average score (%)
Logistic Regression (LR)	0.81
Linear Discriminant Analysis (LDA)	0.81
Quadratic Discriminant Analysis (QDA)	0.32
k-Nearest neighbors Classifier (KNN)	0.68
Random Forest Classifier (RFC)	0.71
Gaussian NB (GNB)	0.61

925
926
927 **Table 3.** The results of cross-validation with six different classification algorithms
928 against the type of products in three classes (a group of [CH_4, C_2H_4, C_2H_5OH], [CH_4,
929 C_2H_5OH, C_3H_7OH], and [CH_4, C_2H_5OH])
930

	Average score (%)		
ML Algorithms	**CH_4, C_2H_4, C_2H_5OH**	**CH_4, C_2H_5OH, C_3H_7OH**	**CH_4, C_2H_5OH**
Logistic Regression (LR)	0.39	0.20	0.80
Linear Discriminant Analysis (LDA)	0.52	0.15	0.93
Quadratic Discriminant Analysis (QDA)	0.43	0.51	0.60
***k*-Nearest neighbors Classifier (KNN)**	0.48	0.65	0.73
Random Forest Classifier (RFC)	0.70	0.71	1
Gaussian NB (GNB)	0.52	0.45	0.26

931
932
933
934
935
936
937
938
939
940
941
942
943
944

945

946 **Table 4.** Evaluation of predictive algorithms for applied potential, faradic efficiency and current density

947

	Statistical Technique	**Bagging Regression (BR)**		**Random Forest Regression (RFR)**		**Gradient Boosting Regression (GBR)**		**Extra Trees Regression (ETR)**	
Features	n/a	Training	Test	Training	Test	Training	Test	Training	Test
Applied Potential(V vs RHE)	MSE	1.17E-03	6.15E-03	5.30E-04	3.18E-03	8.14E-04	2.96E-03	6.31E-06	2.66E-03
	R2	0.97	0.88	0.98	0.94	0.98	0.94	0.99	0.94
Current Density(mA.cm^{-2})	MSE	9.30E-04	7.06E-03	9.76E-04	5.87E-03	1.11E-03	5.42E-03	4.45E-08	4.82E-03
	R2	0.98	0.88	0.98	0.90	0.98	0.91	0.99	0.91
Faradaic Efficiency(%)	MSE	2.44E-03	7.30E-03	1.86E-03	6.22E-03	7.17E-03	8.35E-03	2.56E-31	5.38E-03
Overall	R2	0.98	0.95	0.98	0.96	0.95	0.94	1	0.96

948

949

950

951

952

953