ACS Sustainable Chemistry & Engineering

pubs.acs.org/journal/ascecg

Research Article

Evaluation of Machine Learning Models on Electrochemical CO_2 Reduction Using Human Curated Datasets

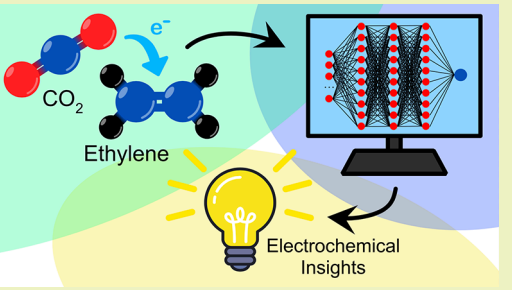
Brianna R. Farris, Tevin Niang-Trost, Michael S. Branicky, and Kevin C. Leonard*

Cite This: *ACS Sustainable Chem. Eng.* 2022, 10, 10934−10944

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Machine learning holds the potential to be a powerful tool to aid in designing catalytic and sustainable chemical systems. However, it is important for experimental researchers to understand the capabilities of different machine learning models when trained on experimental data. In this work, we trained three different machine learning algorithms (decision tree, random forest, and multilayer perceptron) with a hand-curated dataset of 127 reaction conditions for electrocatalytic CO_2 reduction on heterogeneous catalysts in aqueous electrolytes. The input to the machine learning models were the experimental conditions, and we posed four separate outputs to each of these machine learning algorithms: (1) if the number of proton-coupled electron transfer events was greater than two, (2) if carbon−carbon coupling occurred, (3) if ethylene was the major product, and (4) major product prediction. We observed that with a dataset of this size, all three machine learning models could achieve accuracies between 0.7 and 0.8 for the three binary classification problems (1, 2, and 3). Also, the shallow learning decision tree and random forest models performed equal to or better than the deep learning multilayer perceptron models. In the multiclass classification problem (i.e., predicting the product) the accuracy for all models decreased, with the random forest model producing the highest accuracy of 0.6. Analysis of the models showed that machine learning can independently arrive at conclusions that are well-known in the literature, e.g., that Cu is an important catalyst for producing high-carbon content products, and discern more-complicated patterns, with respect to feature importance.



KEYWORDS: *Machine Learning, CO_2 Conversion, Artificial Intelligence, Electrochemistry, Artificial Neural Network*

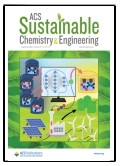
INTRODUCTION

In designing catalytic and sustainable chemical systems, there are an extensive number of process variables (e.g., temperatures, pressures, solvents, catalytic centers, supports, promoters, reactor configurations), and the interplay between these variables is vastly complex. Traditionally, two approaches have been taken for catalyst and reactor design: (1) an Edisonian/empirical/screening approach that relies heavily on experimentation and discovery, and (2) a fundamental approach where mechanistic insights are used (often complemented by computational chemistry) to guide experimental research. Recently, there is an open question regarding how machine learning can be used to augment chemical and catalytic research.

The field of machine learning is based on designing software algorithms that learn from data, discern patterns, and make predictions.[1] This powerful data-driven approach is becoming increasingly popular in many fields including medicine,[2] material science,[3] energy,[4] and engineering,[5−7] because of the increasing availability and improvements in machine learning tools and datasets.[8] While still not used extensively in the field of catalysis and sustainability, groups are starting to use machine learning algorithms such as artificial neural networks to help predict catalyst performance.[9−14] In addition, it is becoming increasingly popular to use a combination of density functional theory (DFT) calculations and machine learning algorithms to accelerate the search for new catalysts.[8,15−18]

Taking a complementary approach of using machine learning trained on computational datasets, in this work, we aim to discover what type of insights off-the-shelf machine learning algorithms trained on experimental datasets can reveal about the electrochemical reduction of CO_2 on heterogeneous surfaces in aqueous electrolytes. While there is some precedent in the literature for machine learning algorithms to be used on electrocatalytic data, this area has yet to be fully explored. For example, Palkovits et al.[10] were able to predict overpotentials of the oxygen evolution reaction using artificial neural networks, support vector regression, and K-nearest neighbor regression models. Their input features consisted of elemental composition of four different catalyst metals with a large dataset of over 6000 experimental data points found in the literature. Their results confirmed that simple machine learning regression algorithms

Received: May 17, 2022
Revised: July 28, 2022
Published: August 10, 2022



can predict overpotential on the oxygen evolution reaction, which has a single reaction product and typically well-defined experimental conditions. Here, we chose to investigate how classification machine learning algorithms would fare on electrochemical CO_2 reduction, because the reaction pathway is more complex and it is popular in sustainable chemistry research. Specifically, the electrochemical reduction of CO_2 is seen as a popular route for enabling the electrification of the chemical industry and for utilizing a ubiquitous greenhouse gas pollutant.

There are several pathways for electrochemical CO_2 reduction on heterogeneous surfaces, depending on how strongly the electrocatalysts adsorbs key CO_2 intermediates. For example, it is well-known that CO_2 can be electrochemically reduced via $2e^-/2H^+$ transfers with high selectivity and efficiency to either CO (on electrocatalysts that strongly adsorbed CO_2 intermediates, e.g., Au)[19,20] or formate (on electrocatalysts that weakly adsorbed CO_2 intermediates, e.g., Sn[21,22] or Bi).[23,24]

However, copper electrocatalysts have long been known to produce several C2 (i.e., two-carbon) products in aqueous media.[25−28] This C−C coupling occurs uniquely on Cu surfaces, because the CO adsorption energy sits at an optimum, which promotes the formation of C2 products.[29−34] While there is evidence that the local pH changes caused by basic cations in the electrolyte can change the selectivity,[35] it is still not clear what all of the variables are that effect the major product formation in the electrochemical CO_2 reduction reaction.

To demonstrate the ability of machine learning algorithms to augment catalytic research, we proposed four classification questions to three different machine learning models trained on a custom-built dataset collected from literature. One of the road blocks of using machine learning is obtaining datasets of sufficient size so that the algorithms can detect patterns. Thus, to determine if machine learning could be a useful tool to the individual catalytic researcher, we restricted ourselves to a dataset of the size that is typically collected by a traditional literature search. Thus, our dataset consisted of 127 examples taken from 106 manuscripts (see the Supporting Information for a full list of manuscripts). We used off-the-shelf shallow learning (e.g., decision tree and random forest) and deep learning (artificial neural network) algorithms to determine if the algorithms could accurately predict (1) if the system would undergo a two-electron transfer or more than a two-electron transfer, (2) if the major product was a multicarbon product or a single carbon product, (3) if the major product was (or was not) ethylene, and (4) the major product.

METHODS

Compiling the Dataset and Data Processing. The dataset used in this study was human-curated and was limited to a size that could be reasonably obtained by a single researcher. To reduce the time used to search for data in the literature, and to mimic the actual literature search processes of a traditional catalysis researcher, review papers containing tabular electrochemical CO_2 reduction data were utilized. The compiled dataset consists of 127 data points of different catalysts and reaction conditions. The full dataset is provided in the Supporting Information.

Feature Selection. When building machine learning algorithms, it is important to decide which features (inputs) to include in the model. We included features that we found to have a high likelihood of authors including that information across multiple manuscripts. In this study, the features extracted from the literature include catalyst metal, dopant, structure, electrolyte, form, potential, and product. Some features that may impact electrochemical CO_2 reduction, or that were not widely or uniformly reported in the literature, were not included in the database. For example, for electrolytes commonly reported within the literature, the pH is not commonly given.

Many of the features used in this dataset were categorical. For example, we assigned the feature form to indicate if the catalyst was either a metallic, oxide, nitride, or chalcogenide. The structure feature was composed of 28 different types of catalyst structures including polycrystalline, nanoparticles, and films. All categorical features were label-encoded. However, the features catalyst metal and dopant were encoded using the atomic number of the metal to give physical meaning to the encoding. Numerical features such as applied potential were entered as reported using the reversible hydrogen electrode (RHE) as the potential reference or were converted to RHE using pH. If the literature did not disclose a pH value, it was estimated using the electrolyte used. A full list of all categorical features and their label encodings are included in Tables S1, S2, and S3 in the Supporting Information.

Machine Learning Packages. The Jupyter notebook framework and Python programming language were used for all of the machine learning studies. The decision tree and random forest algorithms were implemented using the scikit-learn machine learning libraries. The Keras library, which is a user-friendly wrapper for the Tensorflow, was used for all artificial neural networks. All source code for this study can be found in the Supporting Information.

Model Validation. Model validation was performed by splitting the data with 80% in the training set and 20% in the testing set. To further ensure the accuracy scores for the machine learning algorithms, 5-fold cross validation was performed. The algorithms used the scikit-learn `kfold` function with 5 splits to create training and testing sets that incorporate the entire dataset. The scores are calculated for each validation set and an average is taken to give the cross-validation score. For decision trees and random forests, the max depth parameter was varied and for neural networks the number of hidden layers and nodes were varied. The models were then run to determine the parameters that produced the highest accuracy using F-1 score, precision, and recall.

RESULTS AND DISCUSSION

Dataset and Machine Learning Models. Electrochemical CO_2 reduction can undergo several different reaction pathways (a subset is shown in Table 1). There are three binary classification questions relevant to electrochemical CO_2 conversion research:

Table 1. Possible Electrochemical CO_2 Reduction Reactions Pathways along with Whether the Reaction Pathway Undergoes More than Two Electron Transfers, Undergoes C−C Coupling, or Produces Ethylene

product	2+ electrons	multicarbon	ethylene
$CO_2 + 2H^+ + 2e^- \rightarrow HCOOH$	N	N	N
$CO_2 + 1H^+ + 2e^- \rightarrow HCOO^-$	N	N	N
$CO_2 + 2H^+ + 2e^- \rightarrow CO + H_2O$	N	N	N
$CO_2 + 8H^+ + 8e^- \rightarrow CH_4 + 2H_2O$	Y	N	N
$2CO_2 + 7H^+ + 8e^- \rightarrow C_2H_3OO^- + 2H_2O$	Y	Y	N
$2\ CO_2 + 10H^+ + 10e^- \rightarrow CH_3CHO + 3H_2O$	Y	Y	N
$2\ CO_2 + 12H^+ + 12e^- \rightarrow C_2H_4 + 4H_2O$	Y	Y	Y
$2\ CO_2 + 12H^+ + 12e^- \rightarrow C_2H_5OH + 3H_2O$	Y	Y	N
$2\ CO_2 + 14H^+ + 14e^- \rightarrow C_2H_6 + 4H_2O$	Y	Y	N
$2\ CO_2 + 18H^+ + 18e^- \rightarrow C_3H_7OH + 5H_2O$	Y	Y	N

https://doi.org/10.1021/acssuschemeng.2c02941
ACS Sustainable Chem. Eng. 2022, 10, 10934−10944

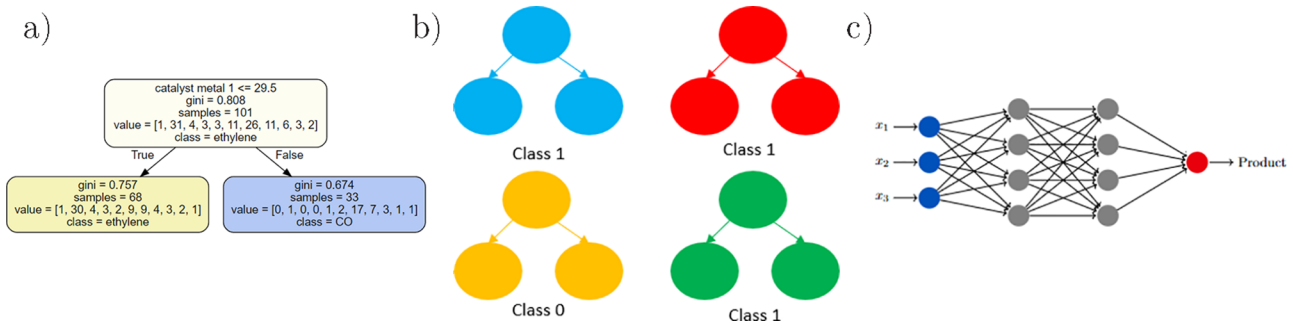


Figure 1. Representations of machine learning models investigated: (a) decision trees, (b) random forest, and (c) multilayer perceptron.

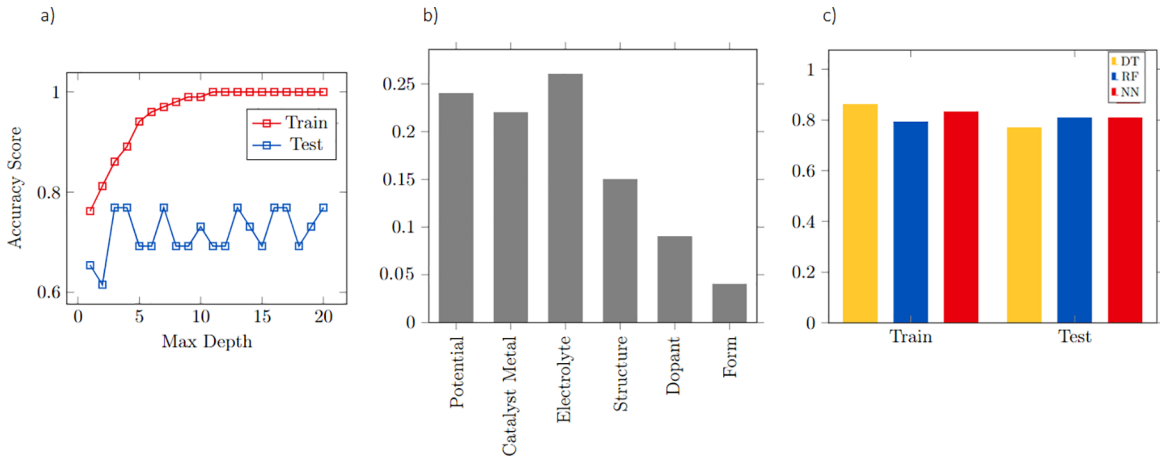


Figure 2. Machine learning results predicting the number of electrons transferred in the electrochemical reduction of CO_2: (a) accuracy versus maximum depth for the decision tree; (b) feature importance obtained from the random forest; and (c) accuracy comparison for decision tree, random forest, and multilayer perceptron.

(1) Will specific catalyst/reaction conditions undergo a simple two-electron reduction or undergo more than a two-electron reduction?
(2) Is the major product single carbon or multicarbon?
(3) Is the major product ethylene or not ethylene?

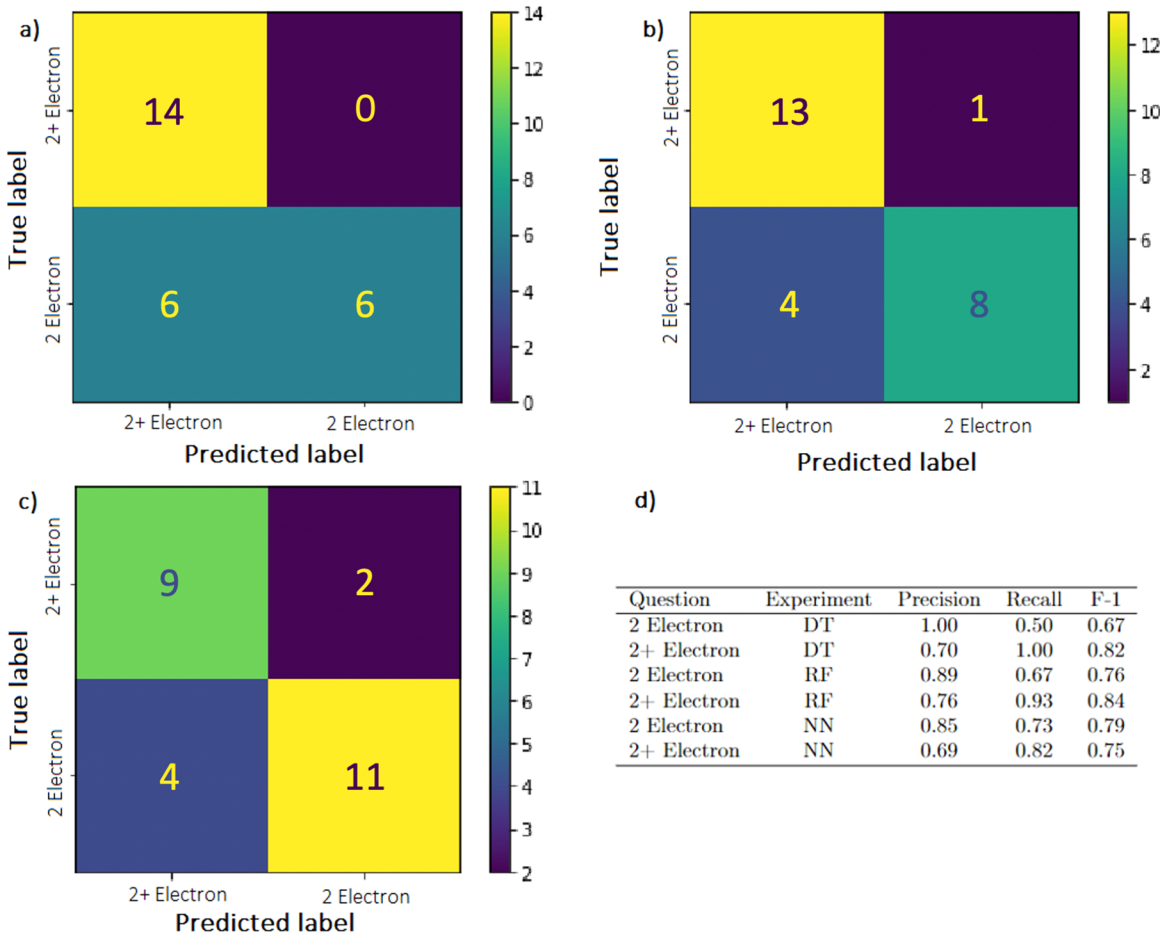
For these binary classification questions, the dataset is relatively well balanced. Out of the 127 reaction conditions in the dataset, there are 59 products that result from a two-electron transfer versus 68 that result from more-than-a-two-electron transfer, 55 multicarbon products versus 72 single carbon products, and 40 that have ethylene as the major product and 87 that do not have ethylene.

Figure 1 shows the models used to ascertain the ability of off-the-shelf machine learning algorithms to aid in electrocatalytic design. We started with decision trees (Figure 1a), which are supervised learning algorithms that split datasets based on descriptive features that enable the resulting sub-datasets to be as pure as possible (with respect to output label). We also investigated random forest algorithms (Figure 1b), which consist of a large number of individual decision trees that operate as an ensemble, where each individual decision tree is trained on a subset of the dataset, and the final output is a majority vote among all the trees. Lastly, we used multilayer perceptron artificial neural networks, which are nonlinear regression or "deep learning" models inspired by the human brain. All of these models were evaluated using simple and user-friendly Python packages (sci-kit learn for the decision tree and random forest models and Keras for the artificial neural network) and all four questions were evaluated for all models.

Numbers of Electrons Transferred. In this set of machine learning experiments, the experimental and catalyst conditions were used as the features, and the label was whether the major reaction pathway was a two-electron transfer or more than a two-electron transfer. A summary of the machine learning results are shown in Figures 2 and 3.

Even though decision trees are relatively simple machine learning algorithms, the fact that they are able to split data based on minimizing the gini impurity (i.e., separate the data into bins, which makes each bin as pure as possible), we observed several correlations by analyzing decision trees. Supporting Information Figure S1a shows an example tree with a depth of two. Intuitively, one may expect that the root node (i.e., the feature that would separate the data best) for achieving more than two proton-coupled electron transfer events would be if the catalyst metal was copper. However, this model showed that splitting the data on the applied potential first actually separated the data the most. For example, of the 127 total samples, 52 samples had an applied potential more negative than −0.878 V vs RHE. Of these samples with highly negative applied potentials, 41 were able to produce products that had more than two proton-coupled electron transfer events. Contrastingly, of the 49 samples that had potentials less negative than −0.879 V vs RHE, only 13 were able to achieve reactions with more than two proton-coupled electron transfer events. Moreover, of those 41 samples where the potential was more negative than −0.879 V vs RHE, 40 used potassium-based electrolytes. This indicates that applying a potential more negative than −0.879 V with potassium-based

https://doi.org/10.1021/acssuschemeng.2c02941
ACS Sustainable Chem. Eng. 2022, 10, 10934−10944



Question	Experiment	Precision	Recall	F-1
2 Electron	DT	1.00	0.50	0.67
2+ Electron	DT	0.70	1.00	0.82
2 Electron	RF	0.89	0.67	0.76
2+ Electron	RF	0.76	0.93	0.84
2 Electron	NN	0.85	0.73	0.79
2+ Electron	NN	0.69	0.82	0.75

Figure 3. Confusion matrix for predicting the number of electrons transferred in the electrochemical reduction of CO_2: (a) decision tree, (b) random forest, and (c) multilayer perceptron. (d) summary of accuracy, precision, and recall. DT = Decision Tree, RF = Random Forest, NN = Neural Network.

electrolytes is important for achieving more than two proton-coupled electron transfer events.

A drawback to decision trees is that it is easy to obtain an overfit model by simply increasing the depth of the tree (i.e., allowing the tree to keep splitting the data). To avoid overfitting, the training and testing set accuracy scores were plotted versus the max depth shown in Figure 2a. The max depth of the final decision tree was chosen based on when the testing set accuracy no longer increased as the training set accuracy increased. Therefore, the max depth for the more than two-electron predicting tree was three layers (see Figure S1b in the Supporting Information). This tree was able to produce an accuracy score of 0.77 on the testing set. This increased accuracy was obtained by adding in the catalyst metal in the third layer. A 5-fold cross-validation was performed over the entire dataset to ensure the training set did not overly influence the output accuracy. The cross-validation score for this dataset was determined to be 0.78, demonstrating the quality of the model. The precision, recall, and confusion matrix were also calculated for this decision tree and are shown in Figure 3. From the precision, recall, and confusion matrix, it can be determined that all of the samples in the testing set that actually produced more than two proton-coupled electron transfer events were correctly predicted by the model. However, if the sample actually only produced a two-proton coupled electron transfer, the model's accuracy was 0.5. There were six total two-proton coupled electron transfer events that were incorrectly categorized as more than two proton-coupled electron transfers.

An advantage of decision trees is that you can determine at which decision on the tree the training set was incorrectly classified (see the Supporting Information). For example, this decision tree has three layers and eight child leaves. Many of these child leaves have very low gini impurity and, thus, are good for accurate prediction. For example, there were 40 samples that had an applied potential more negative than −0.879 V vs RHE and used 1 of the first 13 electrolytes and had 1 of the first 5 structures (see Tables S1 and S2 in the Supporting Information for electrolyte and structure encodings). Of these 40 samples, 38 underwent more than two-electron transfer events for the major product. In addition, there were 18 samples that had potentials less negative than −0.801 V vs RHE and used a catalyst that was Pd or heavier. All of these samples produced two-electron transfer products. However, there was one child leaf with a gini impurity of only 0.475. These were samples that had potentials less negative than −0.879 V vs RHE, a catalyst metal that was copper (or lighter), and were either metallic or a chalcogenide. Of these 19 samples, 11 produced more than two-electron transfer products, and 7 produced two-electron transfer products. Thus, Cu catalysts at potentials less negative than −0.979 V vs RHE will sometimes undergo more than two electron transfers and will sometimes undergo only two electron transfers. Interestingly, 3 of the 6 data points in the testing set that were incorrectly predicted also occurred at this leaf node.

https://doi.org/10.1021/acssuschemeng.2c02941
ACS Sustainable Chem. Eng. 2022, 10, 10934−10944

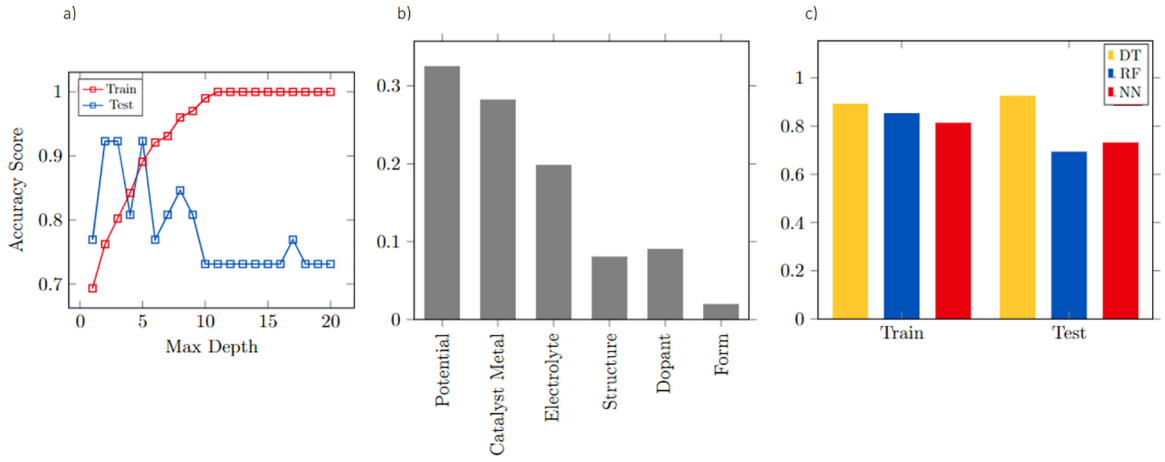


Figure 4. Machine learning results predicting whether the major product underwent C−C coupling in the electrochemical reduction of CO_2: (a) accuracy versus maximum depth for the decision tree; (b) feature importance obtained from the random forest; and (c) accuracy comparison for decision tree, random forest, and multilayer perceptron.

These were samples that used Cu catalysts, but with a low applied potential, and the major product was only a two-electron transfer product.

The second algorithm used to predict whether the reaction undergoes more than two proton-coupled electron transfer events was random forest. Random forests use multiple trees and take the most popular prediction from the ensemble of trees as their final output. Traditionally, training multiple trees and different subsets of the dataset can produce models with higher accuracy. Here, the random forest algorithm for the proposed questions was trained with 100 decision trees over a range of maximum depths (see the Supporting Information).

For the two-electron prediction, the max depth given by the training and testing set accuracy was determined to be 1 (see the Supporting Information). The random forest with a max depth of 1 produced an accuracy score of 0.8. The confusion matrix shows that the random forest algorithm predicted 13 out of the 14 greater than two proton-coupled electron transfer events correctly, while the two-proton coupled electron transfer was predicted in 8 out of the 12 instances.

Similar to the decision tree, all data points in the testing set that used Cu (or lighter) catalysts were predicted to have more than two electron transfers. However, all four of the testing set samples that were incorrectly classified by the model were samples that used copper, but had major products that were only two-electron transfers. Furthermore, the sample in the testing set that was misclassified by the model as a two-proton coupled electron transfer used silver. All other samples in the testing set with catalysts heavier that copper were classified as a two-proton coupled electron transfer.

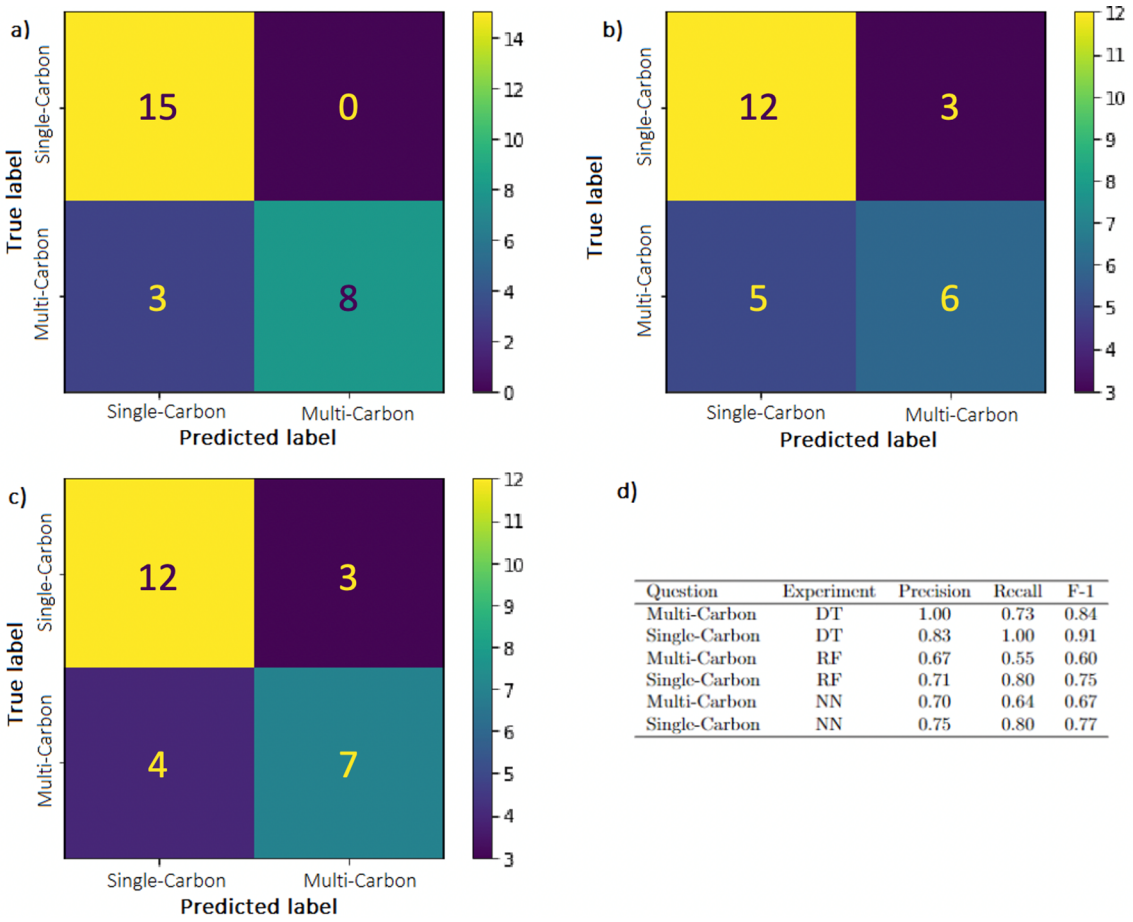
Comparing these results to the decision tree confusion matrix, the random forest algorithm slightly increases the precision of the more than two proton-coupled electron transfer events and recall of the two-proton coupled electron transfer events. However, with this increase comes a slight decrease in the precision and recall of the two-proton coupled electron transfer events and the greater than two proton-coupled electron transfer events, respectively. Similar to the decision trees, cross-validation was performed over the entire dataset for the random forest algorithm, and the cross-validation accuracy was found to be 0.77.

Even though examining splits on a single decision tree can give some information on which features affect the outcome the most, random forest models can give more quantitative feature importance data, because they train over multiple decision trees. As seen in Figure 2b, the higher the feature importance, the more the model used the feature to make decisions. The random forest algorithm with the max depth of 1 indicated that the electrolyte and potential were the two most important features, closely followed by catalyst metal. This matches what we saw from the single decision tree and verifies that the applied potential and electrolyte are important features that affect the products of electrochemical CO_2 reduction.

The last algorithm used for prediction was an artificial neural network. Artificial neural networks are popular because they can produce highly accurate models on complex datasets. However, the disadvantage is that they often require large datasets for training, and it is difficult to obtain insights such as feature importance. To predict the number of proton-coupled electron transfer events, the neural network was tuned using the Keras hyperband tuner to obtain the best number of layers, number of nodes, and learning rate for the data. The tuner gave a model with 5 layers and 1 output layer containing 80, 176, 96, 176, 160, and 1 nodes. The best learning rate was found to be 0.001. The number of nodes determined by the Keras tuner to give the highest accuracy was large for the size of the dataset. Generally smaller amounts of nodes in neural networks are considered better. However, the Keras tuner removes the variability of manually determining the best number of layers and nodes. Interestingly, the neural network model performed with the same accuracy as the decision tree with these parameters, giving a model with an accuracy score of 0.77. The 5-fold cross-validation accuracy was calculated to be 0.71.

The confusion matrix for the neural network shows the two proton-coupled electron transfer event was correctly predicted 11 times, while 4 samples were misclassified as greater than two proton-coupled electron transfers. The more than two proton-coupled electron transfer samples were correctly predicted 9 times, with only 2 samples being incorrectly labeled as a two proton-coupled electron transfer. Since neural networks can be described as a "black box" model, the structure of the network does not give insights to how the predictions are made. However, the prediction of the testing set data may be able to give us clues on how it predicts. In the testing set, 11 out of the 26 samples had catalysts heavier than copper. All but 1 sample with catalysts heavier than copper were labeled as a two-proton-

https://doi.org/10.1021/acssuschemeng.2c02941
ACS Sustainable Chem. Eng. 2022, 10, 10934−10944



d)

Question	Experiment	Precision	Recall	F-1
Multi-Carbon	DT	1.00	0.73	0.84
Single-Carbon	DT	0.83	1.00	0.91
Multi-Carbon	RF	0.67	0.55	0.60
Single-Carbon	RF	0.71	0.80	0.75
Multi-Carbon	NN	0.70	0.64	0.67
Single-Carbon	NN	0.75	0.80	0.77

Figure 5. Confusion matrix for predicting the whether a C−C coupling occurred in the electrochemical reduction of CO_2: (a) decision tree. (b) random forest, and (c) multilayer perceptron. (d) summary of accuracy, precision, and recall. DT = Decision Tree, RF = Random Forest, NN = Neural Network.

coupled electron transfer by the model. The single sample that was not classified as a two proton-coupled electron transfer was 1 of the 4 misclassified as greater than two-proton-coupled electron transfer in the testing set. Moreover, out of the samples heavier than copper in the testing set, only 1 sample with a Pt catalyst was incorrectly classified as a two-proton-coupled electron transfer.

Multicarbon Prediction. The next question that we asked the same three algorithms to predict is whether the product was a multicarbon or single-carbon product (see Figures 4 and 5). The difference between this question and the more than two proton-coupled electron transfer reactions is that products such as methane and methanol undergo more than two electron transfer events, but do not undergo C−C coupling.

The decision tree for the multicarbon prediction had a root node of catalyst metal atomic number of less than 29.5. This is an interesting finding, because we used a hierarchical label encoding for catalyst metal based on atomic number. Thus, when the model predicts that multicarbon products are formed with an atomic number less than 29.5, this implies that the catalyst metal is copper (or lighter). The model also predicted that, for catalysts with an atomic number larger than 29.5, the classification was single carbon. For those catalysts whose atomic number was less than 29.5 (i.e., copper or lighter), further improvement could be made by again looking if the applied potential was highly negative. Similar to above, the training and testing set accuracy was plotted and the max depth of the tree that gave the highest accuracy was determined to be 5. The final decision tree can be found in the Supporting Information.

The decision tree gave an accuracy score of 0.88 with high precision and recall for both the single carbon and multicarbon predictions. The confusion matrix indicates that all 15 of the single-carbon instances in the testing set were correctly predicted as single carbon. For the multicarbon predictions, 8 out of the 11 were correctly predicted as multicarbon products with 3 instances being incorrectly labeled as single-carbon products. By comparing the 3 misclassified data points in the training set to the decision tree, we can see where the algorithm misclassified. The first sample incorrectly labeled as single carbon had a high atomic number of 78, indicating Pt as the catalyst. Coincidentally, this sample was the only one in the database to have Pt as its catalyst. Therefore, there was no training data for the model to learn from to correctly classify this as a multicarbon product. This indicates that the Pt catalyst may be incorrectly classified in other models, because of the lack of representative training data. The second data point incorrectly classified was a Cu catalyst with a potential of −0.85 V that produced a multicarbon product. It was misclassified as single carbon due to its "low potential" and its electrolyte being K_2SO_4. Interestingly, the child leaf that this data point was split into had a gini of 0, meaning that all the testing data in that leaf were single-carbon samples. The last misclassified sample in the testing set was classified as single carbon, because the catalyst

https://doi.org/10.1021/acssuschemeng.2c02941
ACS Sustainable Chem. Eng. 2022, 10, 10934−10944

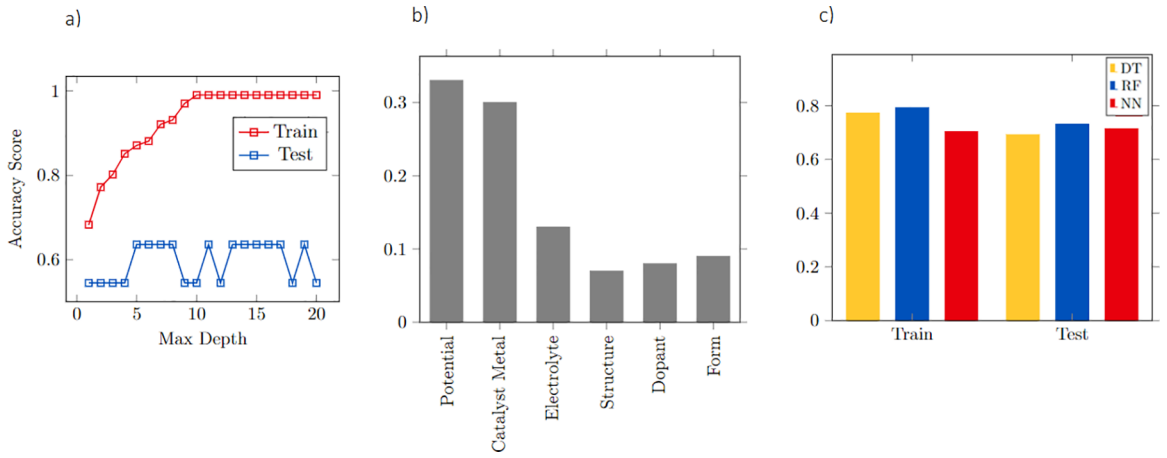
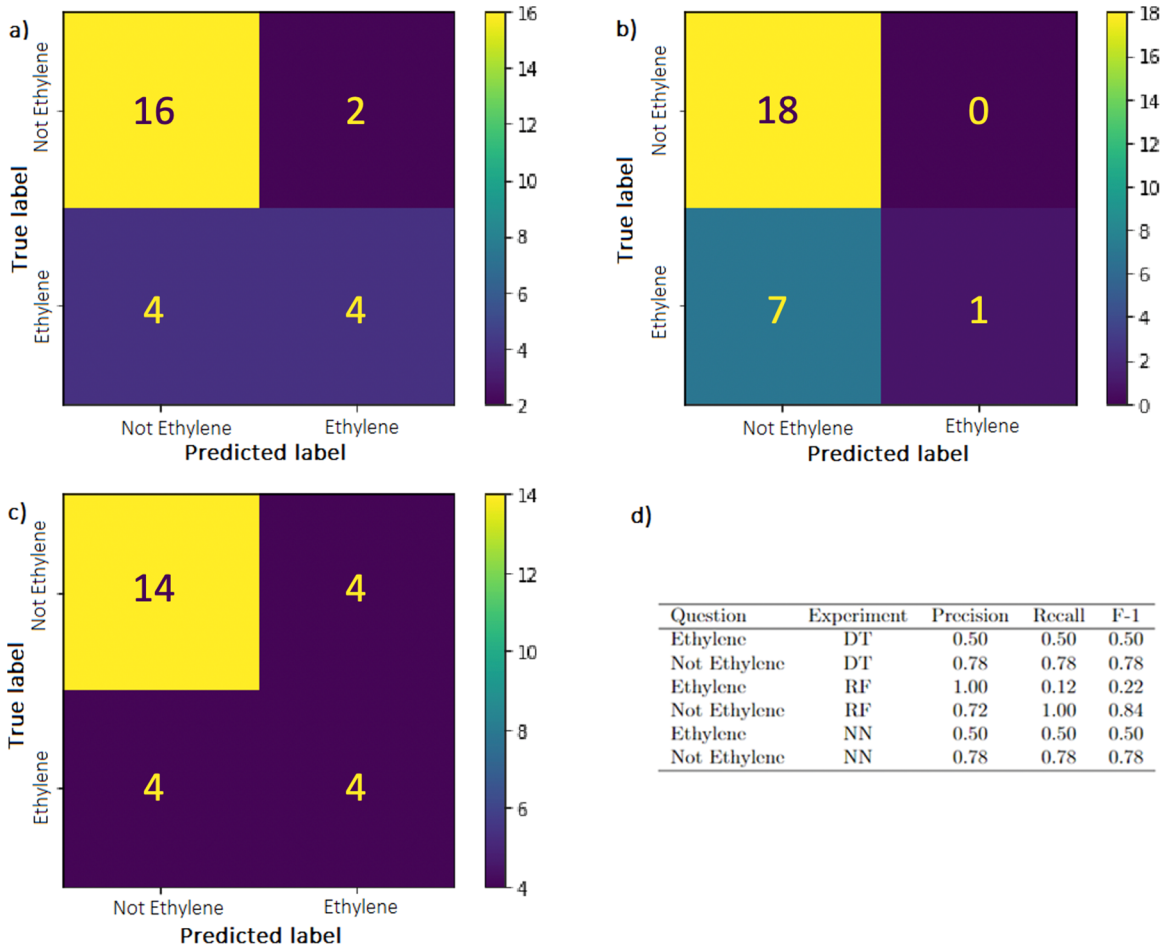


Figure 6. Machine learning results predicting whether the major product was ethylene or not ethylene: (a) accuracy versus maximum depth for the decision tree; (b) feature importance obtained from the random forest; and (c) accuracy comparison for decision tree, random forest, and multilayer perceptron.



Question	Experiment	Precision	Recall	F-1
Ethylene	DT	0.50	0.50	0.50
Not Ethylene	DT	0.78	0.78	0.78
Ethylene	RF	1.00	0.12	0.22
Not Ethylene	RF	0.72	1.00	0.84
Ethylene	NN	0.50	0.50	0.50
Not Ethylene	NN	0.78	0.78	0.78

Figure 7. Confusion matrix for predicting whether ethylene production occurred in the electrochemical reduction of CO_2: (a) decision tree, (b) random forest, and (c) multilayer perceptron. (d) summary of accuracy, precision, and recall. DT = Decision Tree, RF = Random Forest, NN = Neural Network.

metal was nickel. Interestingly, the training set data there contained one data point with a nickel catalyst that had a potential more negative than −0.876 V that produced methane, a single-carbon product. Therefore, the model predicted the testing sample with a nickel catalyst to be single carbon as well. Cross-validation was performed over the entire dataset and the accuracy was found to be 0.76.

The random forest algorithm was trained over 100 trees and a range of maximum depths (see the Supporting Information). The max depth found to be optimal for multicarbon prediction was a max depth of 2. With a max depth of 2, the accuracy score

https://doi.org/10.1021/acssuschemeng.2c02941
ACS Sustainable Chem. Eng. 2022, 10, 10934−10944

resulting from the algorithm was 0.69. The low accuracy score can be attributed to the low precision and recall of the multicarbon class predictions. The confusion matrix shows that 12 out of the 15 single-carbon testing samples were predicted correctly, while only 6 out of the 11 multicarbon instances in the testing set were predicted correctly as multicarbon products. The confusion matrix along with the larger precision, recall and f1-score indicate that the random forest algorithm was better at predicting single-carbon products than multicarbon products.

By looking at the testing set predictions, a trend is seen with the 3 misclassified single-carbon predictions. All three used copper catalysts with potentials more negative than −1.0 V and were classified as multicarbon by the random forest algorithm. In addition, the three misclassified catalysts produced methane instead of a multicarbon product. Out of the 5 samples misclassified as single carbon, 4 of the samples had copper as the catalyst metal, with the fifth sample using silver. The copper catalysts that were misclassified all had potentials less negative than −0.86 V and used either $KHCO_3$ or KOH as their electrolytes. The cross-validation score calculated was higher with an accuracy of 0.79, demonstrating the importance of having good representation between the training and testing sets.

The feature importance was found for the multicarbon prediction using random forest algorithm with a max depth of 2. The feature with the highest importance was potential, with an importance of 0.325. The second highest was catalyst metal with 0.28. This is similar to what we saw from the single decision tree with the root node of catalyst metal, indicating that it is an important feature for the multicarbon production from electrochemical CO_2 reduction.

As was done with the previous model, the neural network was tuned using the Keras hyperband tuner to obtain the best number of layers, number of nodes, and learning rate for the data. The tuner gave a model with 5 layers and 1 output layer containing 104, 32, 80, 64, 144, and 1 nodes. The best learning rate was found to be 0.001. The neural network model with these parameters gave a model with an accuracy score of 0.73. The neural network had the lowest cross validation accuracy out of the multicarbon predictions with a values of 0.72.

The confusion matrix shows that the neural network predictions were similar to the random forest predictions with 12 out of the 15 single carbon testing samples being predicted correctly. Out of the 3 misclassified single-carbon products, 2 of the samples used copper catalysts and 1 used palladium. The neural network performed slightly better than the random forest algorithm with the multicarbon predictions, with 7 out of the 11 instances being predicted correctly. The samples misclassified as single carbon included 3 using copper catalysts and 1 using the only a platinum sample in the database. From the copper catalysts incorrectly labeled as single carbon, 2 contained dopants of silver and chlorine.

Ethylene Prediction. In this set, we asked the models to evaluate the binary classification problem of whether the major product was ethylene or not ethylene (see Figures 6 and 7). The decision tree gave an accuracy of 0.77 with a max depth of 2. The decision tree had a root node of catalyst metal atomic number less than 29.5, again indicating that the catalyst metal is copper (or lighter). Both child nodes stemming from this root node were classified as nonethylene products. However, for catalyst metals larger than copper, none of the training samples included an ethylene sample, making that a pure child leaf. The child node with catalyst metals copper and lighter contained a split of 39 non ethylene samples and 32 ethylene samples. This set could then be further split if the potential was more negative than ca. −0.88 V vs RHE.

The confusion matrix shows that 16 out of the 18 nonethylene samples were correctly predicted, while 4 out of the 8 ethylene products were correctly predicted. The two samples in the testing set inaccurately classified as ethylene were copper and nickel catalysts with potentials greater than −0.876 V. The 3 out of 4 samples in the testing set that did produce ethylene but were misclassified as nonethylene products all used copper as the catalyst with potentials less negative than −0.876. The last sample that produced ethylene but was misclassified as nonethylene used Pt as the catalyst causing the misclassification.

Again, the random forest algorithm was trained over 100 trees and a range of maximum depths. The random forest algorithm had an optimal max depth of 2, resulting in an accuracy of 0.73. The confusion matrix shows that the algorithm predicted mostly nonethylene samples. It correctly predicted all of the non-ethylene samples. However, it predicted 7 out of the 8 ethylene samples as nonethylene.

Out of the 7 misclassified ethylene samples, 6 of them used copper catalysts with a range of potentials from −0.5 V to −1.38 V vs RHE. Since copper catalysts can produce a wide range of products, it is possible with a dataset of this size that there is not enough data for the random forest algorithm to correctly determine when ethylene would be made. The last missclassified sample was the single platinum catalyst in the database with a potential of −0.8 V vs RHE.

The feature importance for the ethylene predictions with the entire database was found using the random forest algorithm. The top feature for splitting the data was a potential at 0.32, which was closely followed by catalyst metal with 0.30 importance. This matches what we saw from the single decision tree's root node of catalyst metal and verifies that the applied potential and catalyst metal are important features that affects the production of ethylene from electrochemical CO_2 reduction.

As was done with the previous models, the neural network was tuned using the Keras hyperband tuner to obtain the best number of layers, number of nodes, and learning rate for the data. The neural network consisted of 5 layers and an output layer consisting of 256, 128, 192, 120, 200, and 1 node, respectively. The optimum learning rate was found to be 0.01. The neural network model with these parameters gave a model with an accuracy score of 0.69. Out of the three algorithms for ethylene predictions, the neural networks gave the worst accuracy score. From the confusion matrix, it is seen that the nonethylene samples were predicted correctly in 14 out of the 18 instances. The ethylene samples were predicted correctly half of the time with 4 samples out of the 8 correctly predicted as ethylene. The cross-validation score was found to be 0.69.

Interestingly, a recurring theme among all of the machine learning models is that applied potential and electrolyte all have high feature importance. To visualize this importance, a comparison of the potential and electrolyte was plotted in Figure 8 for the ethylene data using copper catalysts. Only reactions using copper catalysts were plotted since most of the ethylene samples in the dataset used copper as the metal. By plotting the potential versus electrolyte, a trend for the usage of $KHCO_3$ was observed. Using $KHCO_3$ as the electrolyte between −1.00 V and −1.50 V vs RHE, ethylene was more likely to be the major product. However, with KOH as the electrolyte, ethylene was produced preferably at less-positive potentials. This optimum range of applied potential is an example of an

https://doi.org/10.1021/acssuschemeng.2c02941
ACS Sustainable Chem. Eng. 2022, 10, 10934−10944

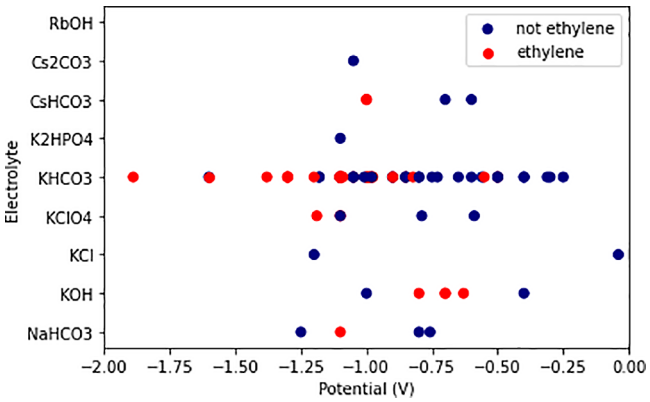


Figure 8. Comparison of potential and electrolyte values for the ethylene and nonethylene data.

interesting finding that machine learning can yield in the field of electrocatalysis. With even larger datasets available, machine learning may indicate more relationships between applied potential and electrolyte on product production.

Product Prediction. The last proposed question was whether the machine learning algorithms can perform multiclass predictions and predict the exact product of the reaction (see Figures 9 and 10). The potential products included acetaldehyde, acetate, ethylene, ethanol, ethane, propanol, methane, CO, formate, formic acid, hydrogen, and methanol.

First, a decision tree was used for the multiclass prediction. The product prediction tree was found to have a max depth of 3 (see Figure S4 in the Supporting Information). The root node was used if the catalyst metal was less than 29.5 (Cu or lighter). If the root node was true, then the model would predict ethylene with 68 samples containing 30 ethylene occurrences. If the root node was false, the model would predict CO with 33 samples containing 17 CO occurrences. A major issue with the decision tree's multiclass predictions was that the tree would not have enough leaves to predict each product. With 3 layers, the tree was only able to predict 4 out of the 12 possible products in the database. However, at higher depths, the tree became overfit and the accuracy decreased. In addition, with the size of the dataset, the products were not evenly distributed through the training and testing sets. Therefore, the precision and recall were only calculated for the 4 predicted products, leading to an accuracy score of 0.5 with low precision and recall scores overall. The cross-validation score was lower than the accuracy score with a value of 0.40.

The four products that the decision trees predicted were ethylene, methane, CO, and formate. From the confusion matrix, seen in Figure S8 in the Supporting Information, the decision tree correctly predicted ethylene 7 times, CO 5 times, and formate once. However, it incorrectly predicted ethylene 3 times, methane 3 times, and CO 7 times. Ethylene and CO were more likely to be predicted by the model, since they were the two products most commonly found in the training set.

Next, random forest was used to predict the products. The max depth was found to be 4, with an accuracy of 0.62. However, similar to the decision tree out of the 5 products in the testing set, the model only predicted 4 of the products, as seen in Figure 10. The random forest model was able to accurately separate if the reaction conditions would produce ethylene or CO with these products having accuracy scores of 0.82 (9 out of 11) and 0.86 (6 out of 7), respectively. However, the model had difficulty with the other products, because of the lack of representation of these samples in the dataset. The cross-validation score was found to be 0.49. Similar to the binary classification predictions for multicarbon and ethylene, the potential was found to be the most important feature for product predictions.

Lastly, the neural network model was used for product prediction. The neural network contained 5 layers and 1 output layer with 16, 64, 128, 64, 32, and 21 nodes, respectively. The model was trained over 1000 epochs giving an accuracy of 0.46 with only 12 out of the 26 testing samples being accurately predicted. Out of the 8 possible products in the testing set, only 5 products were predicted. However, only 4 of the predicted products including ethylene, formate, CO, and hydrogen were accurately predicted (see Figure S9 in the Supporting Information). This led to low precision and recall values for the predictions. Unlike the decision tree and random forest algorithms, the class predicted the best by the neural network was CO, followed by ethylene. Cross-validation was performed for the neural network, resulting in an accuracy of 0.42.

Overall, the main challenge of multiclass prediction was the amount and distribution of the data. For all of the algorithms, there was not enough data to evenly distribute the classes to the training and testing sets. Thus the training set data was not representative of the testing sets leading to low accuracy scores.

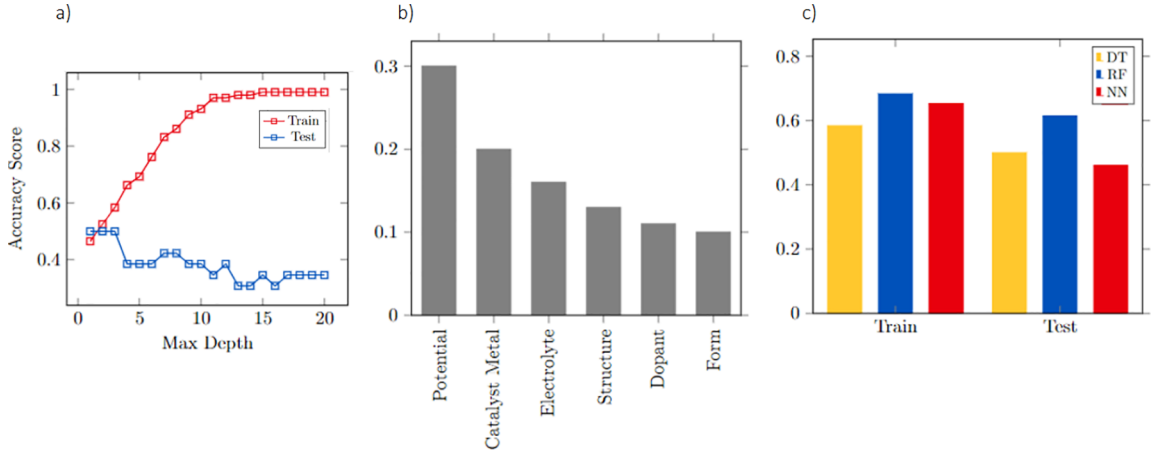


Figure 9. Machine learning models for product prediction: (a) accuracy versus maximum depth for the decision tree; (b) feature importance obtained from the random forest; and (c) accuracy comparison for decision tree, random forest, and multilayer perceptron.

https://doi.org/10.1021/acssuschemeng.2c02941
ACS Sustainable Chem. Eng. 2022, 10, 10934−10944

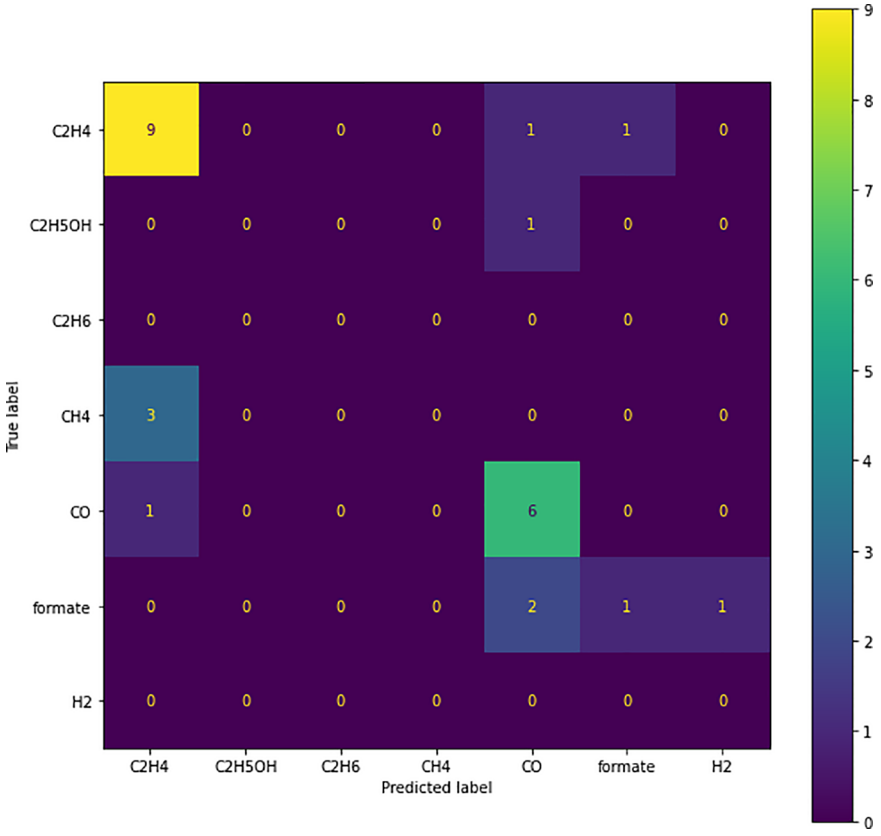


Figure 10. Example of confusion matrix for the product prediction from the random forest algorithm.

Similarly, an effective data collection method is crucial for increasing the size of the dataset. The manual process of human-curated datasets is limited to the time frame of the curator.

CONCLUSIONS

In this work, we showed that, on hand-curated datasets of the size that a single researcher could obtain, machine learning can provide relatively high accuracies with binary classification questions. In addition, we show that deep learning multilayer perceptron models do not necessarily outperform simple shallow learning models such as decision trees and random forests. Moreover, these shallow learning models can provide additional insights such as the feature importance with the random forest. We found that these machine learning models can independently arrive at some conclusions already well-established in the literature, e.g., that copper is an important catalyst for producing high-carbon content products. In addition, these models can show less obvious insights such as the importance of the applied potential. Looking more closely at the applied potential may help researchers in this area examine their systems. Lastly, to get product prediction (multiclass classification), a dataset of ca. 100 samples is not sufficiently large when there are 12 possible products. To build models with this capability, it will be imperative to develop automatic data mining of the literature to build datasets significantly larger than what a single researcher can obtain in a reasonable time.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssuschemeng.2c02941.

Additional tables and graphics including the full decisions trees, data encoding, max depth determination graphs, product confusion matrices, the full reference list (PDF)

Source code for all models including the required packages and preprocessing code to prepare the data for use in the model (PDF)

AUTHOR INFORMATION

Corresponding Author

Kevin C. Leonard − *Department of Chemical & Petroleum Engineering, The University of Kansas, Lawrence, Kansas 66045, United States; Center for Environmentally Beneficial Catalysis, The University of Kansas, Lawrence, Kansas 66047, United States;* orcid.org/0000-0002-0172-3150; Email: kcleonard@ku.edu

Authors

Brianna R. Farris − *Department of Chemical & Petroleum Engineering, The University of Kansas, Lawrence, Kansas 66045, United States; Center for Environmentally Beneficial Catalysis, The University of Kansas, Lawrence, Kansas 66047, United States*

Tevin Niang-Trost − *Department of Electrical Engineering & Computer Science, The University of Kansas, Lawrence, Kansas 66045, United States*

Michael S. Branicky − *Department of Electrical Engineering & Computer Science, The University of Kansas, Lawrence, Kansas 66045, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acssuschemeng.2c02941

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by a the U.S. National Science Foundation Research Traineeship (NRT) grant through Award No. DGE-1922649. The authors would also like to thank Josh Meckstroth for assistance with some of the images.

■ REFERENCES

(1) Jordan, M. I.; Mitchell, T. M. Machine Learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255−260.
(2) Khan, J.; Wei, J. S.; Ringner, M.; Saal, L. H.; Ladanyi, M.; Westermann, F.; Berthold, F.; Schwab, M.; Antonescu, C. R.; Peterson, C.; Meltzer, P. S. Classification and Diagnostic Prediction of Cancers using Gene Expression Profiling and Artificial Neural Networks. *Nature Medicine* **2001**, *7*, 673−679.
(3) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Computational Materials* **2017**, *3*, 1−13.
(4) Liu, Y.; Esan, O. C.; Pan, Z.; An, L. Machine learning for advanced energy materials. *Energy and AI* **2021**, *3*, 100049.
(5) Peng, H.; Ling, X. Optimal design approach for the plate-fin heat exchangers using neural networks cooperated with genetic algorithms. *Applied Thermal Engineering* **2008**, *28*, 642−650.
(6) Rorke, D. C. S.; Suinyuy, T. N.; Gueguim Kana, E. B. Microwave-assisted chemical pre-treatment of waste sorghum leaves: Process optimization and development of an intelligent model for determination of volatile compound fractions. *Bioresour. Technol.* **2017**, *224*, 590−600.
(7) Romo, J. E.; Miller, K. C.; Sundsted, T. L.; Job, A. L.; Hoo, K. A.; Wettstein, S. G. The Effect of Solvent Polarity on Autocatalytic Furfural Production Confirmed by Multivariate Statistical Analysis. *ChemCatChem.* **2019**, *11*, 4715−4719.
(8) Goldsmith, B. R.; Esterhuizen, J.; Liu, J.-X.; Bartel, C. J.; Sutton, C. Machine Learning for Heterogeneous Catalyst Design and Discovery. *AIChE J.* **2018**, *64*, 2311−2323.
(9) Kite, S.; Hattori, T.; Murakami, Y. Estimation of catalytic performance by neural network - product distribution in oxidative dehydrogenation of ethylbenzene. *Appl. Catal. A* **1994**, 173−178.
(10) Palkovits, R.; Palkovits, S. Using Artificial Intelligence To Forecast Water Oxidation Catalysts. *ACS Catal.* **2019**, *9*, 8383−8387.
(11) Serra, J. M.; Corma, A.; Chica, A.; Argente, E.; Botti, V. Can artificial neural networks help the experimentation in catalysis? *Catal. Today* **2003**, *81*, 393−403.
(12) Chowdhury, A. J.; Yang, W.; Abdelfatah, K. E.; Zare, M.; Heyden, A.; Terejanu, G. A. A Multiple Filter Based Neural Network Approach to the Extrapolation of Adsorption Energies on Metal Surfaces for Catalysis Applications. *J. Chem. Theory Comput.* **2020**, *16*, 1105−1114.
(13) Holena, M.; Baerns, M. Feedforward neural networks in catalysis: A tool for the approximation of the dependency of yield on catalyst composition, and for knowledge extraction. *Catal. Today* **2003**, *81*, 485−494.
(14) Yazdizadeh, M.; Jafari Nasr, M. R.; Safekordi, A. A new catalyst for the production of furfural from bagasse. *RSC Adv.* **2016**, *6*, 55778−55785.
(15) Jinnouchi, R.; Asahi, R. Predicting Catalytic Activity of Nanoparticles by a DFT-Aided Machine-Learning Algorithm. The. *J. Phys. Chem. Lett.* **2017**, *8*, 4279−4283.
(16) Zhai, H.; Alexandrova, A. Ensemble-Average Representation of Pt Clusters in Conditions of Catalysis Accessed through GPU Accelerated Deep Neural Network Fitting Global Optimization. *J. Chem. Theory Comput.* **2016**, *12*, 6213−6226.
(17) Deng, C.; Su, Y.; Li, F.; Shen, W.; Chen, Z.; Tang, Q. Understanding activity origin for the oxygen reduction reaction on bi-atom catalysts by DFT studies and machine-learning. *Journal of Materials Chemistry A* **2020**, *8*, 24563−24571.
(18) Toyao, T.; Suzuki, K.; Kikuchi, S.; Takakusagi, S.; Shimizu, K.-i.; Takigawa, I. Toward Effective Utilization of Methane: Machine Learning Prediction of Adsorption Energies on Metal Alloys. *J. Phys. Chem. C* **2018**, *122*, 8315−8326.
(19) Chen, Y.; Li, C. W.; Kanan, M. W. Aqueous CO2 Reduction at Very Low Overpotential on Oxide-Derived Au Nanoparticles. *J. Am. Chem. Soc.* **2012**, *134*, 19969−19972.
(20) Lucio, A. J.; Shaw, S. K. Pyridine and Pyridinium Electrochemistry on Polycrystalline Gold Electrodes and Implications for CO2 Reduction. *J. Phys. Chem. C* **2015**, *119*, 12523−12530.
(21) Chen, Y.; Kanan, M. W. Tin Oxide Dependence of the CO2 Reduction Efficiency on Tin Electrodes and Enhanced Activity for Tin/Tin Oxide Thin-Film Catalysts. *J. Am. Chem. Soc.* **2012**, *134*, 1986−1989.
(22) Zhang, S.; Kang, P.; Meyer, T. J. Nanostructured Tin Catalysts for Selective Electrochemical Reduction of Carbon Dioxide to Formate. *J. Am. Chem. Soc.* **2014**, *136*, 1734−1737.
(23) Medina-Ramos, J.; Pupillo, R. C.; Keane, T. P.; DiMeglio, J. L.; Rosenthal, J. Efficient Conversion of CO2 to CO Using Tin and Other Inexpensive and Easily Prepared Post-Transition Metal Catalysts. *J. Am. Chem. Soc.* **2015**, *137*, 5021−5027.
(24) DiMeglio, J. L.; Rosenthal, J. Selective Conversion of CO2 to CO with High Efficiency Using an Inexpensive Bismuth-Based Electrocatalyst. *J. Am. Chem. Soc.* **2013**, *135*, 8798−8801.
(25) Feng, X.; Jiang, K.; Fan, S.; Kanan, M. W. Grain-Boundary-Dependent CO2 Electroreduction Activity. *J. Am. Chem. Soc.* **2015**, *137*, 4606−4609.
(26) Alves, D. C.; Silva, R.; Voiry, D.; Asefa, T.; Chhowalla, M. Copper nanoparticles stabilized by reduced graphene oxide for CO2 reduction reaction. *Materials for Renewable and Sustainable Energy* **2015**, *4*, 1−7.
(27) Roberts, F. S.; Kuhl, K. P.; Nilsson, A. High Selectivity for Ethylene from Carbon Dioxide Reduction over Copper Nanocube Electrocatalysts. *Angew. Chem., Int. Ed. Engl.* **2015**, *54*, 5179−5182.
(28) Loiudice, A.; Lobaccaro, P.; Kamali, E. A.; Thao, T.; Huang, B. H.; Ager, J. W.; Buonsanti, R. Tailoring Copper Nanocrystals towards C2 Products in Electrochemical CO2 Reduction. *Angew. Chem., Int. Ed. Engl.* **2016**, *55*, 5789−5792.
(29) Clark, E. L.; Hahn, C.; Jaramillo, T. F.; Bell, A. T. Electrochemical CO2 Reduction over Compressively Strained CuAg Surface Alloys with Enhanced Multi-Carbon Oxygenate Selectivity. *J. Am. Chem. Soc.* **2017**, *139*, 15848−15857.
(30) Hori, Y.; Murata, A.; Yoshinami, Y. Adsorption of CO, intermediately formed in electrochemical reduction of CO2, at a copper electrode. *Journal of The Chemical Society, Faraday Transactions* **1991**, *87*, 125−128.
(31) Hori, Y.; Koga, O.; Yamazaki, H.; Matsuo, T. Infrared spectroscopy of adsorbed CO and intermediate species in electrochemical reduction of CO2 to hydrocarbons on a Cu electrode. *Electrochim. Acta* **1995**, *40*, 2617−2622.
(32) Zhang, Y.-J.; Sethuraman, V.; Michalsky, R.; Peterson, A. Competition between CO2 Reduction and H2 Evolution on Transition-Metal Electrocatalysts. *ACS Catal.* **2014**, *4*, 3742−3748.
(33) Genovese, C.; Ampelli, C.; Perathoner, S.; Centi, G. Electrocatalytic conversion of CO2 to liquid fuels using nanocarbon-based electrodes. *Journal of Energy Chemistry* **2013**, *22*, 202−213.
(34) Bushuyev, O. S.; De Luna, P.; Dinh, C. T.; Tao, L.; Saur, G.; van de Lagemaat, J.; Kelley, S. O.; Sargent, E. H. What Should We Make with CO2 and How Can We Make It? *Joule* **2018**, *2*, 825−832.
(35) Resasco, J.; Lum, Y.; Clark, E.; Zeledon, J. Z.; Bell, A. T. Effects of anion identity and concentration on electrochemical reduction of CO2. *ChemElectroChem.* **2018**, *5*, 1064−1072.

https://doi.org/10.1021/acssuschemeng.2c02941
ACS Sustainable Chem. Eng. 2022, 10, 10934−10944