ChemElectroChem

Research Article
doi.org/10.1002/celc.202300647

Chemistry Europe
European Chemical Societies Publishing

www.chemelectrochem.org

Understanding the Performance of (Ni–Fe–Co–Ce)O_x-Based Water Oxidation Catalysts via Explainable Artificial Intelligence Framework

Paul Rossener Regonia*[a] and Christian Mark Pelicano*[b]

Among the most active oxygen evolution reaction (OER) catalysts, mixed metal oxides based on Ni, Fe, and Co metals are recognized as economical yet excellent replacements for RuO_2 and IrO_x. However, tuning and searching for optimal compositions of multi–element–compound electrocatalysts is a big challenge in catalysis research. Conventional materials screening experiments and theoretical simulations are labor–intensive and time–consuming. Machine learning offers a promising paradigm for accelerating electrocatalyst research and simultaneously understanding composition–activity correlation. Herein, we introduce an Explainable AI (XAI) framework for predicting the electrocatalytic performance of OER catalysts. By integrating the robust Random Forest (RF) model for machine learning with the Shapley Additive Explanations (SHAP) method for model explanation, we achieved accurate predictions of the overpotential for various compositions of (Ni–Fe–Co–Ce)O_x catalysts ($R^2 = 0.8221$). More importantly, we obtained valuable insights into how each metal and their interactions influence the overpotential of the catalysts. Our results highlight the versatility of the RF model with SHAP in identifying the optimal composition of (Ni–Fe–Co–Ce)O_x catalysts for electrocatalytic oxygen evolution, showing its potential applicability across various catalyst synthesis methods. Finally, we anticipate that this work will lead to exciting possibilities in designing highly active multi–element compound electrocatalysts with the aid of explainable AI.

Introduction

Developing energy conversion and storage technologies based on sustainable power sources has become an increasingly urgent issue to address the ever-increasing global energy demand.[1–4] Among others, promising alternative energy systems include rechargeable metal–air batteries with high energy densities, and H_2O electrolyzers for producing H_2.[5,6] Conversely, the industrial implementation and efficiencies of these electrochemical technologies are hindered by the slow kinetics of the oxygen evolution reaction (OER). This reaction specifically involves a four–proton and four–electron pathway, which requires a large thermodynamic potential (1.23 V vs RHE), resulting in large energy losses.[7–10] Although noble metals such as IrO_x and RuO_x remain the state–of–the–art materials for OER, their scarcity and high cost hinder their commercial viability.[11] Hence, extensive efforts have been focused on the exploitation of earth–abundant non–precious metals having analogous OER activities with benchmark catalysts.

Most OER electrocatalysts have been traditionally explored via trial–and–error basis and high–throughput screening, which mostly rely on intuition and prior experimental knowledge. Synthesizing one material at a time only slows down the development of new catalysts and is deemed impractical. It is also not likely that all possible parameters or factors affecting the OER activity could be examined. Recent advances in ab initio methods like density functional theory (DFT) have aided researchers in designing and discovering new catalysts at an unprecedented rate.[12,13] However, DFT calculations have been limited in uncovering advanced materials due to confined space/time systems, high computational costs, and sensitive exchange–correlation functionals.[14] Alternatively, the emergence of machine learning (ML) provides a powerful framework to properly address the multi–dimensional and complex processes not only in catalysis but also in other fields.[15–17] With the explosion of data availability in materials science, ML algorithm models can offer an in–depth fundamental insight into the previously unknown relationship or pattern between the materials' structures and their catalytic properties, thereby accelerating the rational design of highly active and stable electrocatalysts.[18] For example, Jiang et al. reported a random forest algorithm to predict the OER activity of hydroxide catalysts under extensive doping space. The model demonstrated a mean relative error of 4.74% in forecasting new experiments.[19] Another study employed a combined high–throughput DFT and ML techniques to develop IrO_2-based electrocatalysts with superior OER activity. By exploiting a neural network language model (NNLM), the relationship

[a] Dr. P. Rossener Regonia
Department of Computer Science, College of Engineering, University of the Philippines Diliman, Philippines
E-mail: pdregonia@up.edu.ph

[b] Dr. C. M. Pelicano
Department of Colloid Chemistry, Max Planck Institute of Colloids and Interfaces, 14476 Potsdam, Germany
E-mail: christianmark.pelicano@mpikg.mpg.de

Supporting information for this article is available on the WWW under https://doi.org/10.1002/celc.202300647

between neighboring atomic environment and the free energies of OER intermediates and formation energies of crystals was established.[20] As a result, a cluster of potential candidates has been found to possess superior OER performance. However, current data–driven ML models often focus on achieving high predictive accuracy at the expense of model interpretability.[21–23] Recently, explainable artificial intelligence (XAI) has been considered a powerful tool in data science as the integration of Shapley values with ML algorithms can help understand the black box of ML predictions.[24] The interpretation of a ML analysis must enable the determination of variables affecting the prediction and provide paths to recognize relevant framework and offer supplementary information that can be adopted to aid decision–making.[24] Esterhuizen et al. applied a principal component analysis in order to yield low–dimensional and explainable electronic–structure descriptors of near–surface alloys and their reactivity origin.[25]

On the basis of these concepts, we present an XAI approach to predicting the overpotential of high–performance OER electrocatalysts based on earth–abundant non–precious metals (Ni–Fe–Co–Ce)O_x). XAI allows post–hoc explanations of complex models such as neural networks, support vectors, and random forests, making it a practical tool for interpreting non–transparent models.[21,23] Our XAI OER model utilized a random forest algorithm that surpassed the predictive performance of previous OER models for (Ni–Fe–Co–Ce)O_x catalysts.[26,27] Shapley additive explanations were generated from the model to gain insights into the significance of its features and its learning process.[23,28] As AI technology becomes more advanced, its inherent complexity poses greater challenges to understanding it. We hope this study will inspire further AI-oriented research in materials science, emphasizing the need for AI explainability and developing unbiased and accessible models.

Experimental Section

Explainable AI in materials science follows four main principles: explanations, meaningfulness, accuracy, and knowledge limits.[21] Explanations provide details about the outcomes of a machine learning (ML) model. These details can be presented graphically or verbally, given that they are meaningful–i.e., the intended audience can easily understand them. The meaningfulness of explanations also depends on their purpose. For instance, when explaining an ML model for predicting material properties, the goal would be to understand how its parameters contribute to the resulting prediction. There is an underlying assumption that the model is accurate; otherwise, the explanations will also be incorrect. However, the accuracy of explanations may change depending on the audience's level of knowledge. An expert materials scientist would comprehend a more detailed explanation than a novice in the field. Thus, balancing between meaningfulness and accuracy depends on the specific audience and situation. But, regardless of the audience's expertise, explanations are still bounded by the knowledge limits of the model's application domain. XAI cannot explain areas where the model was not trained or designed to provide an answer.[21]

In this study, we propose an XAI framework for OER catalyst prediction using machine learning. Figure 1 shows the OER XAI process (right) and applications (left). The process starts with materials synthesis, which includes preparation, synthesis, and characterization of material samples. After collecting sufficient samples, the data goes through preprocessing, visualization, and exploratory analysis–a technique known as Exploratory Data Analysis (EDA). The main objective of EDA is to comprehend the data by checking its quality, summarizing it, and formulating an appropriate model.[29] EDA also identifies outliers, which are typically considered anomalies in materials synthesis data, and are

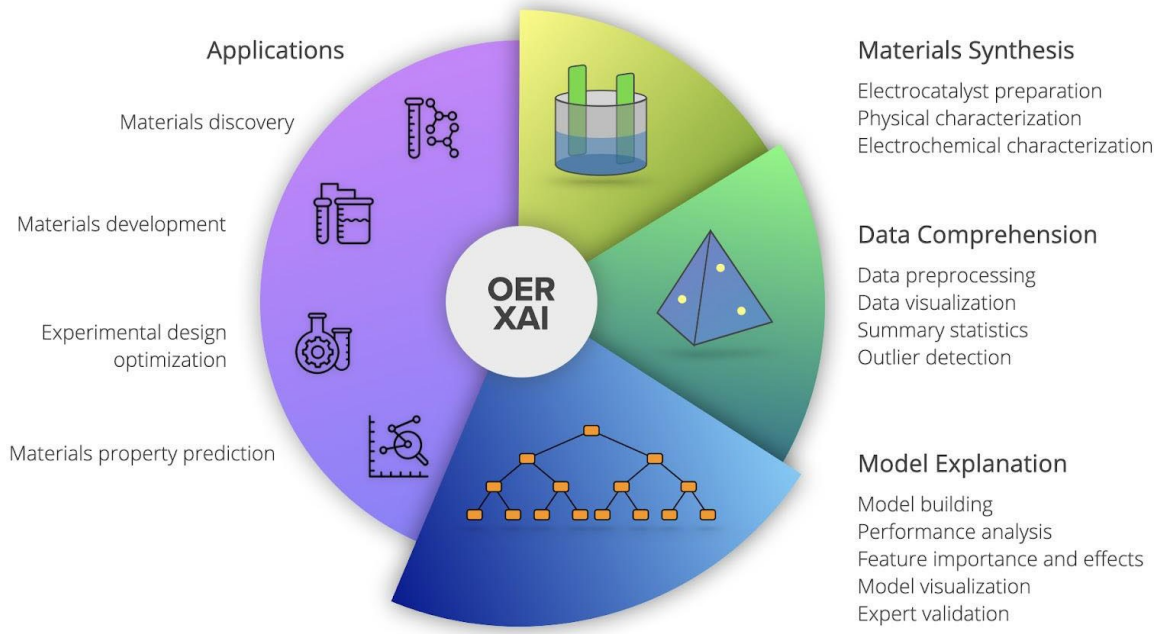

Figure 1. Explainable AI (XAI) framework for oxygen–evolution reaction (OER) modeling.

thus removed from the data set. EDA is an integral part of XAI since understanding the data leads to a better understanding of the model. Following EDA is building and explaining the model, which involves the machine learning pipeline and the generation of meaningful explanations for the ML model.[21,23]

The OER XAI process can be applied in various materials science contexts, such as predictive modeling of OER catalysts, optimizing synthesis processes, developing advanced materials, and–ultimately–discovering novel materials. Consequently, the outcomes of these applications will inform succeeding efforts in explaining AI-driven OER models. It is worth noting that this cyclic process (rather than a linear approach) is adopted in our proposed framework, which signifies an iterative approach to materials synthesis, analysis, modeling, and comprehension, leading to research and industrial applications.

Materials Synthesis and Data Comprehension

Electrocatalyst Data: (Ni–Fe–Co–Ce)O_x composition space was collected from previous studies on Ce-rich family of active OER catalysts.[30,31] The original data set contains 5456 samples of various Ni–Fe–Co–Ce)O_x catalysts spaced at 3.33 at% composition steps. Each catalyst composition was synthesized using inkjet printing, and their electrocatalytic performance was measured using the overpotential (OVP) value at a current density of 10 mA/cm^2.[31]

Exploratory Data Analysis: Our goal in this study is to build an OER XAI model that can predict the overpotential (target) based on the electrocatalyst composition (features). This model should both have high accuracy and a significant degree of explainability. As a precursor to model building, EDA was performed to understand better the underlying characteristics, patterns, and relationships of the data set. EDA involves checking the quality of the data, summarizing it using descriptive statistics, and detecting outliers. In the context of machine learning, the main objectives of EDA are data description followed by model formulation.[29] EDA commonly begins with data preprocessing. In our gathered data set, some samples were excluded due to poor data quality, resulting in difficulty extracting the overpotential.[31] Removing the low-quality data reduced the data set to 5413 samples, with Ni%, Fe%, Co%, and Ce% as the feature variables and OVP as the target variable. Figure 2 shows the 3D quaternary visualization of our data set, highlighting the sample with the lowest OVP.

The next step in EDA is data analysis using descriptive statistics. This step involves calculating statistical measures such as the variables' mean and standard deviation, their skewness and frequency distribution, and the correlation between the features and the target variable.[29] Identifying strong correlations between variables is important in machine learning. Correlation analysis aids in selecting important features while eliminating redundant ones, thereby effectively reducing the dimensionality of the data.[32] Outlier analysis is the final step in our EDA process. Outliers in the data set are samples that deviate significantly from the rest of the data.[29,33] In machine learning, removing outliers is crucial in preventing potential noises in data from influencing the model, which could otherwise lead to biased or inaccurate predictions.[34] However, in materials science studies, it is common for novel materials to exhibit outstanding characteristics, which may be considered outliers from a statistical perspective. Thus, it is important to carefully consider the nature of outliers before deciding to remove them.

Model Building and Explanation

Artificial intelligence is an emerging data–driven approach to materials development and discovery.[22,35,36] Its applications are virtually limitless, and its versatile and agile infrastructure accelerates the trajectory of materials science–more specifically, materials informatics–to new heights. Much of recent progress in materials research employing AI techniques has been centered around supervised machine learning. The primary goal in this domain is to use experimental data to predict material composi-

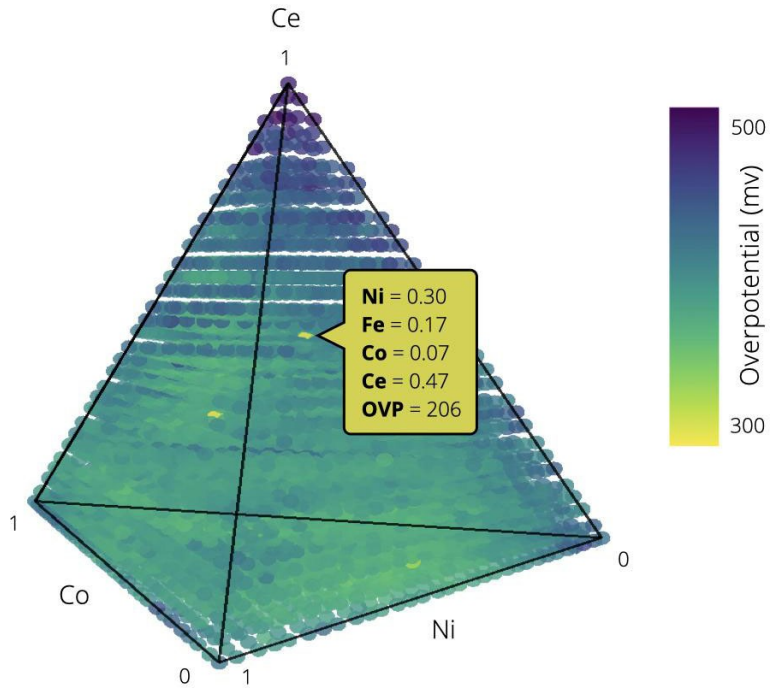

Figure 2. 3D quaternary visualization of (Ni–Fe–Co–Ce)O_x compositions and their overpotential. The catalyst sample with Ni% = 0.30, Fe% = 0.17, Co% = 0.07, and Ce% = 0.47 yielded the lowest overpotential of 206 mV at 10 mA/cm^2.

tions, structures, and properties.[13,16,18,35,36] However, the black–box nature of ML models remains a challenge when it comes to understanding, interpreting, and explaining the underlying mechanisms of such models.[23] In most cases, achieving accurate predictions is just as crucial as comprehending the model's decision process to generate these predictions. In fact, a high level of accuracy does not necessarily indicate that the model is free from bias. Explainable AI/ML addresses this issue by employing AI techniques specially designed to explain the internal mechanisms of ML models. Adopting an XAI framework in AI-supported materials studies is the next step toward building unbiased predictive models.[23] Supervised machine learning has been applied in several renewable energy materials studies,[18] particularly in investigating the oxygen evolution reaction of electrocatalysts.[26,27] Similar to this study, other studies utilized machine learning in predicting the overpotential of (Ni–Fe–Co–Ce)O_x catalysts. Several machine learning models were tested, and support vector regression (SVR) and random forest regression (RFR) emerged as the most notable.[26,27] However, these studies focused on the predictive performance of the ML models. To expand on this research, we incorporate XAI into the OER prediction model to enhance its interpretability by explaining the key features that shape its predictions.

Random Forest Regression: Random forest is an ensemble learning technique that constructs multiple decision trees and makes predictions via a simple majority vote. While traditional decision trees evaluate all features when splitting nodes, random forests restrict their evaluation to a random subset of features. This additional layer of randomness prevents the model from relying too heavily on the training data, making it robust against overfitting.[37] Random Forest regression is a variant of random forest that deals with continuous target variables. RFR models are trained on several hyperparameters, including the number of trees (n_estimators), the number of features to evaluate when splitting (max_features), the maximum depth of trees (max_depth), the minimum number of samples to split (min_samples_split) and to be at a leaf node (min_samples_leaf), and the bootstrapping method (bootstrap).[37,38]

To implement RFR, we used the Scikit–learn package in Python, which has a *RandomForestRegressor* class designed for training RFR models.[38] First, the data set is split into a 70:30 training and testing ratio. This strategy ensures that the model is evaluated on unseen data; otherwise, its performance measure may be biased. A cross–validation technique is performed using the training set to tune the model's hyperparameters. Cross–validation allows for model evaluation during the training stage by creating a validation set from the training set, essentially simulating a test set within training. It can reveal whether the trained model is underfitted (high bias) or overfitted (high variance).[39] We used the Scikit–learn *RandomizedSearchCV* class to find the best hyperparameters through randomized cross-validation.[38,39] Table 1 shows the hyperparameter search space for the RFR model. With 5,632 potential hyperparameter combinations, tuning the hyperparameter can be time–consuming and resource–demanding. However, we streamlined the search space through the randomized search, narrowing it down to 3,000 possibilities. Note that this involves a trade-off between optimizing the model and expediting the training process.

Table 1. Hyperparameter search space for random forest optimization through randomized search cross–validation.

N_estimators: {100, 250, 500, 1000}	*min_samples_leaf:* {1, 2, 4, 6}
max_features: {1, 2, 3, 4}	*min_samples_split:* {2, 5, 10, 15}
max_depth: {10, 20, 30, 40, 50, 60, 70, 80, 90, 100, None}	*bootstrap:* {True, False}

Finally, the tuned model is evaluated through the following performance metrics: mean absolute error (MAE), root mean squared error (RMSE), mean relative error (MRE), and R^2 statistic. The MAE, RMSE, and MRE are derived from the residual error, representing the difference between the predicted and actual observed values. MAE uses the absolute value of errors, making it less sensitive to outliers. On the other hand, RMSE penalizes the model for large residuals, which ensures a more accurate prediction even in the presence of outliers.[40] MRE approximates the error as a proportion of the actual values. It is often more useful than absolute error when comparing multiple models.[41] The R^2 statistic evaluates the model's goodness–of–fit but focuses on the proportion of variability in the target variable that can be explained by the features. An R^2 value of 1 suggests that the model has perfectly fit the unseen data set, indicating strong generalization capabilities for the model.[39]

Shapley Additive Explanation (SHAP): SHapley Additive exPlanations (SHAP) is a unified framework for interpreting ML predictive models. It provides local explanations, revealing how individual features contribute to the model's predictions. SHAP is useful for understanding the features' importance, effects (positive or negative), and additive impact.[21,28] To determine the contribution of each feature, the weighted average of all possible differences in including a feature is calculated as:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{i\}}\left(x_{S \cup \{i\}}\right) - f_S(x_S) \right] \quad (1)$$

where ϕ_i is the Shapley value for feature i; S is a subset of all the features F; x_S is the feature values in S; and the predictions between models including and excluding the feature i are compared. Note that calculating the Shapley values requires evaluating the model on all feature subsets S.[23] This study used the open-source Python SHAP package to generate the feature summary and force plots.[42]

Results and Discussion

Data Characteristics

(Ni–Fe–Co–Ce)O_x compounds were synthesized using high-resolution inkjet printing, and their OER performance was measured in terms of overpotential values. The OVP is an informative measure of the relative electrocatalytic activity, representing the additional energy needed to drive a reaction at a practically achievable rate.[31,43] A low OVP indicates greater energy efficiency in the catalytic compound. Figure 2 illustrates the interaction between Ni, Fe, Co, and Ce metals, and their impact on the resulting OVP. The composition of Ni% = 0.30, Fe% = 0.17, Co% = 0.07 and Ce% = 0.47 yielded the lowest overpotential (OVP = 206 mV). Additional details regarding the compositions with the lowest and highest OVP are available in Figures S1 and S2. These figures suggest that attaining a low OVP involves finding a balance between the metal compositions. On the other hand, exceptionally high levels of Ce are likely to increase the OVP.[31] Table 2 summarizes the OVP statistics. The data set displays a mean OVP of 421.1134 mV

Table 2. Overpotential statistics.				
mean = 421.1134	*std* = 14.9694	*skew* = −0.4152	*min* = 206	*max* = 498
correlation to OVP	*Ni %:* −0.5151	*Fe %:* 0.1134	*Co %:* 0.0113	*Ce %:* 0.3904

with a standard deviation of 14.9694. It is slightly skewed to the right, with values ranging from 206 to 498 mV OVP (Figure 3). Among the four metals, Ni % had the highest (albeit negative) correlation with OVP at −0.5151, followed by Ce %. In contrast, Fe % and Co % showed the weakest correlation.

There are six outliers on the lower end of the data set, as indicated in Figure 3. These outliers fall outside 3 standard deviations from the mean (OVP < 376.2051 mV). Figure S3 shows the relative compositions of each outlier. Since lower OVP values are particularly interesting to us as they indicate higher energy efficiency, these outliers were retained during the data preprocessing phase.

Model Optimization

The data set was randomly split into 70 % training (n = 3789) and 30 % test (n = 1624) data. Additionally, the six low OVP outliers were randomly allocated, with four included in the training set and two in the test set. Figure S4 shows a consistent distribution pattern between the training and test sets concerning both feature and target variables. Our data set source employed a combinatorial program to generate the metal compositions. We verified the uniform distribution of compositions for each metal by tallying the number of samples for each composition value, as shown in Figure S5.

A significant advantage of a large training data set is that it improves the model's capacity to learn more accurately. To verify this, we analyzed the learning curve of the model during cross–validation. Figure S6 shows the model's accuracy progression as the training set size increases. While the training accuracy remains relatively consistent, there is a visible improvement in the validation accuracy, reaching its peak at the maximum number of training samples.

During the hyperparameter tuning, 3000 models were generated using random combinations of hyperparameters (as listed in Table 1) and evaluated using 5-fold cross-validation. Figure 4 shows the randomized search cross–validation results. On average, the hyperparameters with the highest cross–validation mean R^2 scores were: *n_estimators* = 250, *max_features* = 2, *max_depth* = None, *min_samples_split* = 2, *min_samples_leaf* = 2, and *bootstrap* = True. However, the best-performing model among all hyperparameter values were: *n_estimators* = 100, *max_features* = 2, *max_depth* = None, *min_samples_split* = 15, *min_samples_leaf* = 4, and *bootstrap* = True. This discrepancy can be attributed to the random nature of the search process, as randomized search does not explore the entire hyperparameter space. Different results might be obtained when employing a different search algorithm, such as grid search.[38]

Model Performance

Following hyperparameter tuning results, a random forest model consisting of 100 decision tree estimators was trained. While random forests are not strictly black box models, it becomes more challenging to understand them as the forests grow. Figure 5 displays one of the decision trees. Visualizing all 100 of them would be daunting, but analyzing each tree would be even more challenging. Thus, there is a need for an XAI approach.

The model's performance was evaluated on the unseen test data set using multiple accuracy and goodness–of–fit metrics. Table 3 presents the performance scores of the base RFR model (i.e., trained using the default hyperparameters in Scikit–learn) and the hyperparameter–tuned RFR model. Hyperparameter tuning notably enhances the model's performance, as evi-

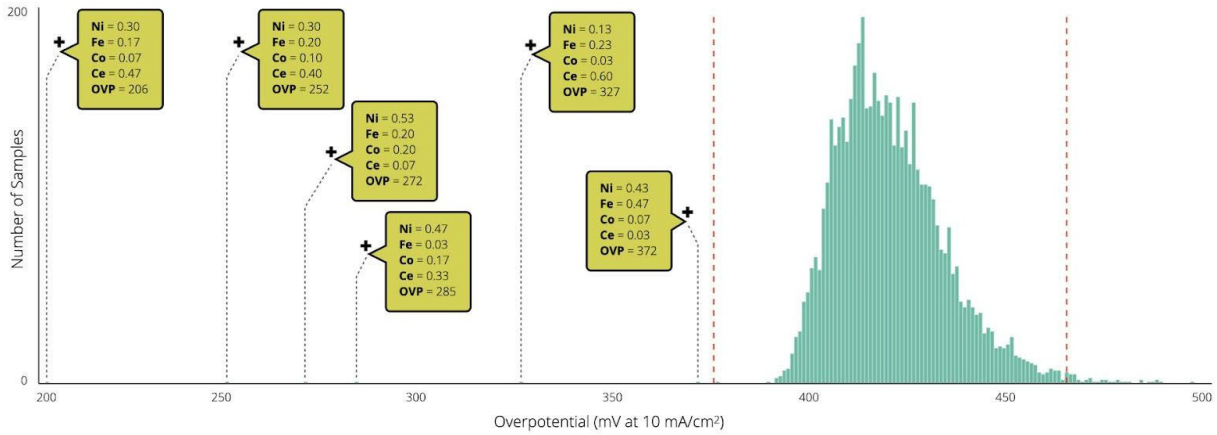

Figure 3. Outliers in the data set.

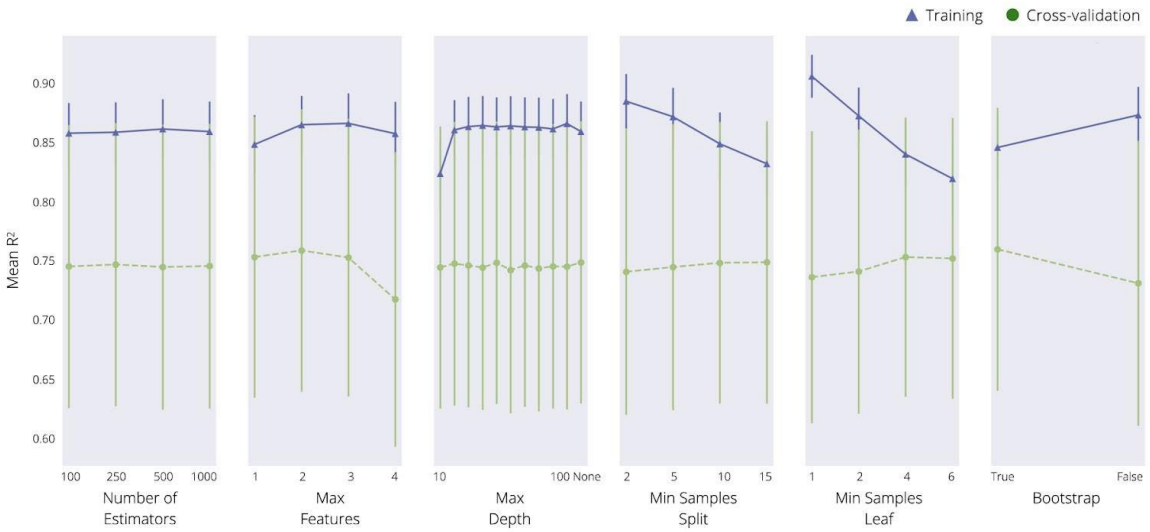

Figure 4. Randomized search cross–validation hyperparameter scores (mean and standard deviation).

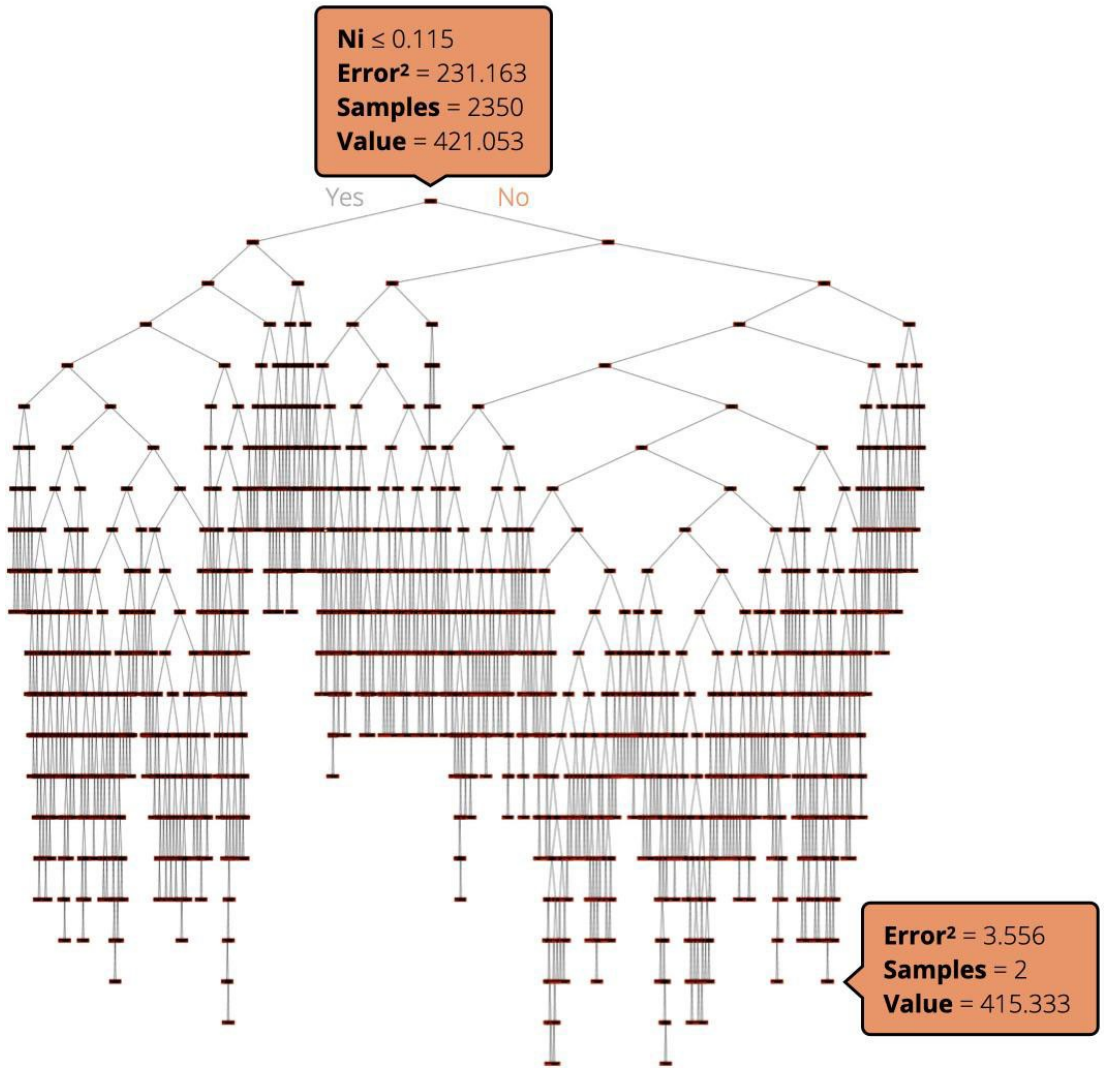

Figure 5. A decision tree in the random forest model.

Table 3. Performance comparison of our base RFR model, our hyperparameter-tuned RFR model, and existing machine learning models for overpotential prediction of (Ni–Fe–Co–Ce)O_x catalysts. The best–performing model for each metric is highlighted.

	MAE	MSE	RMSE	MRE	R^2
Base model	3.9651	40.5839	6.3705	0.9466 %	0.8056
Tuned model	3.9443	37.1339	6.0938	0.9421 %	0.8221
SVR model[21]	–	9.46e–05	0.0097	–	0.7256
RFR + model[22]	–	49.79	7.0562	1.20 %	–

denced by reduced residual errors (MAE, MSE, RMSE, MRE) and increased variance explained (R^2). On average, our model's OVP prediction deviates from the actual values by approximately ± 4 mV (MAE) and ±6 mV when accounting for outliers (RMSE). These errors are considerably small, suggesting that the model is fairly accurate. Our model also effectively captures the underlying patterns and variations in the data set, as indicated by an R-squared value of 82.21 %. The remaining 17.79 % variance may be attributed to the outliers and other unknown factors influencing the OVP. Figure 6 illustrates the residual errors in the model's predictions and demonstrates how outliers impact its performance.

Compared to prior OVP prediction models, our model exhibits notably improved performance. Although our model has a higher residual error than the SVR model, it achieves a higher R^2 score, indicating that it provides a better overall fit to the data. Additionally, despite utilizing the same random forest model with RFR+, we improved the relative errors using a smaller feature set. This result aligns with Occam's razor principle, which favors the simplest model when multiple models produce equally effective results.[39]

Model Explanation

SHAP algorithm was performed to produce explanations for our RFR model. To ensure unbiased explanations of the model, SHAP values were computed separately for the training and test sets. SHAP values represent the feature importance for regression models. They explain how removing a feature impacts the model's prediction.[42] Figure 7 illustrates the individual interactions between metal composition and OVP. Among the metal composition features, Ni exhibited the most substantial influence (i.e., the widest range of SHAP values), with Ce and Co following behind. On the other hand, Fe appeared to have a relatively minimal impact on the OVP (Figure S7).

A feature's SHAP value indicates both its magnitude and direction of impact on the target variable. For instance, a positive SHAP value of 20 indicates that the inclusion of the feature contributes a net increase of 20 to the predicted outcome. Likewise, a negative SHAP value of −10 would reduce the predicted outcome by 10, although its influence is relatively weaker.

Examining the SHAP values of nickel, we observe a consistent trend: as the percentage of Ni increases, the OVP decreases. This is expected since Ni-based compounds possess moderate binding energies (second only to noble metal-based electrocatalysts) based on volcano plots. In turn, a smaller theoretical overpotential could be anticipated towards the OER.[11] Conversely, cerium has the inverse effect: a lower percentage

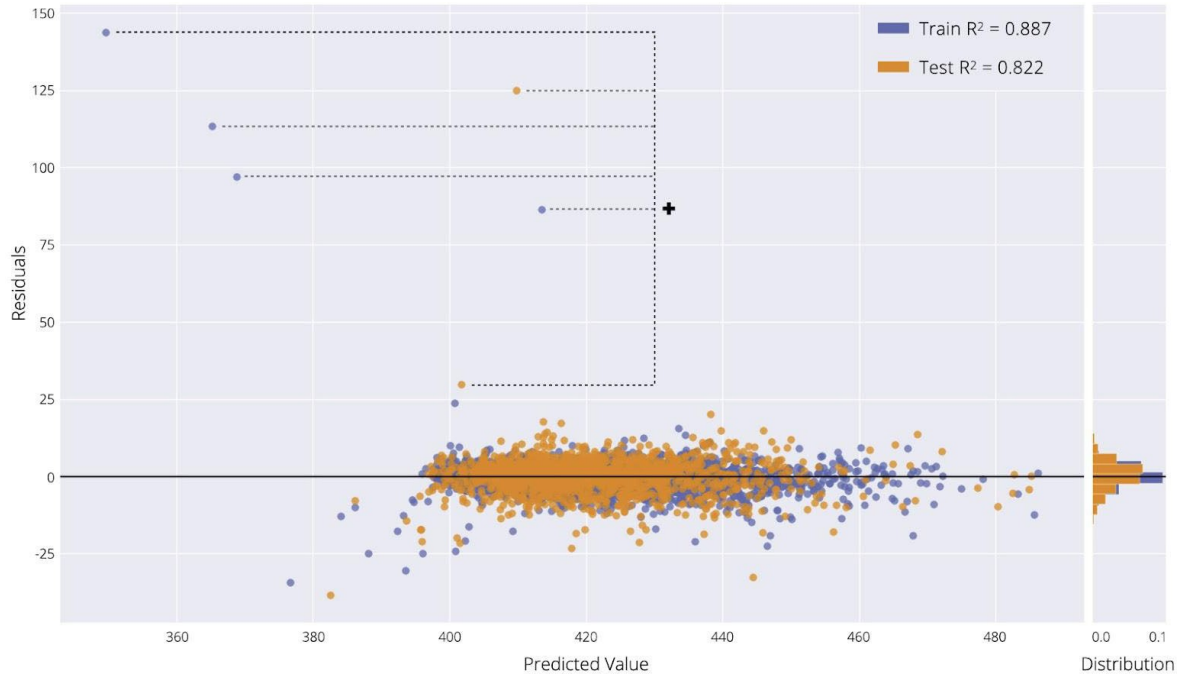

Figure 6. Random forest model prediction, residuals, and outliers.

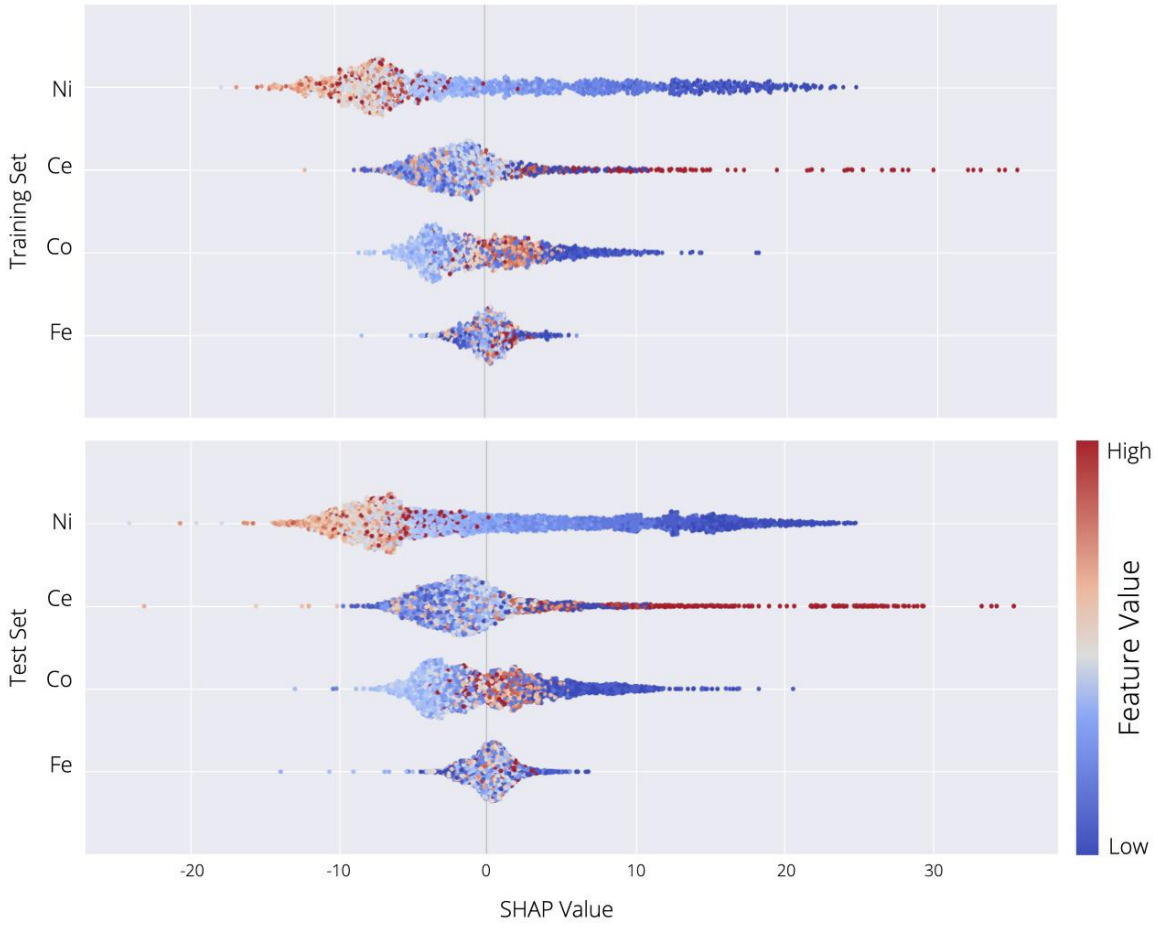

Figure 7. Feature effects on target variable (overpotential). Features are sorted from highest to lowest impact (SHAP value). Note that the SHAP values for both the training and test data sets exhibit similar patterns, indicating that the random–split algorithm produced an even distribution of samples.

of Ce yields lower OVP. Taking into account the inactivity of pure CeOx, a minimal amount of Ce is logical to achieve considerable OER activity.[44] Even more intriguing is the effect of cobalt on the OVP. Reduced Co levels may either increase or decrease the OVP. However, higher Co levels have a comparatively negligible effect. Iron exhibits varied–yet rather minimal–effects on the OVP. However, these effects do not imply that Fe lacks catalytic potential. The SHAP values presented in Figure 7 only reflect the individual contributions of each feature. Hence, it is crucial to comprehend the interactions among features as well. Figure 8 summarizes the strongest interactions between feature pairs. Notably, there is a mutual interaction between Fe and Ni. This finding agrees with previous reports, which established that the interaction of Fe with Ni can dramatically enhance the conductivity of their resulting binary oxide owing to the synergistic effect between the two metal centers.[45] Indeed, mixed Ni–Fe compounds have been regarded as the most promising OER catalysts based on earth-abundant elements for over three decades,[46,47] and their reported activities show little variation regardless of synthesis route.[48–51] Corrigan et al. found that Ni gains a higher oxidation state (Ni^{3+} to Ni^{4+}) via partial electron transfer from Fe^{3+}, which renders Ni species a stronger oxidizing power to oxidize H_2O, thus boosting the OER performance.[52,53] A relatively small amount of Fe (~30%) combined with a substantial proportion of Ni (~30%–60%) results in low OVP, suggesting that this specific combination of Ni and Fe could be a promising area for optimizing electrocatalysts (Figure 8a). There is a tipping point where adding more Ni to Fe decreases the OVP (Figure 8b). In fact, a study by Friebel et al. utilized an operando X-ray absorption spectroscopy to study the influence of increasing Fe content on $Ni_{1-x}Fe_xOOH$ catalysts.[54] They found that above 25% Fe, Fe nucleates as a separate phase based on diverging Fe–O and Ni–O distances, elucidating the absence of activity improvement in excess of this Fe content.

This trend reverses when adding Ni with Ce (Figure 8d). Lastly, reduced Co% and increased Ce% appear to reduce the OVP (Figure 8c). Based on these findings, an appropriate amount of dopant ensures an effective electronic and geometric effect, leading to favorable intermediate binding energies and enhanced OER activities.

The cumulative impact of each feature on the model's prediction is illustrated in Figure 9. In the figure, we observe the baseline prediction, denoted as the SHAP expected value with OVP = 421.33 mV. The figure also reveals the direction and magnitude of each feature's influence. For instance, a composi-

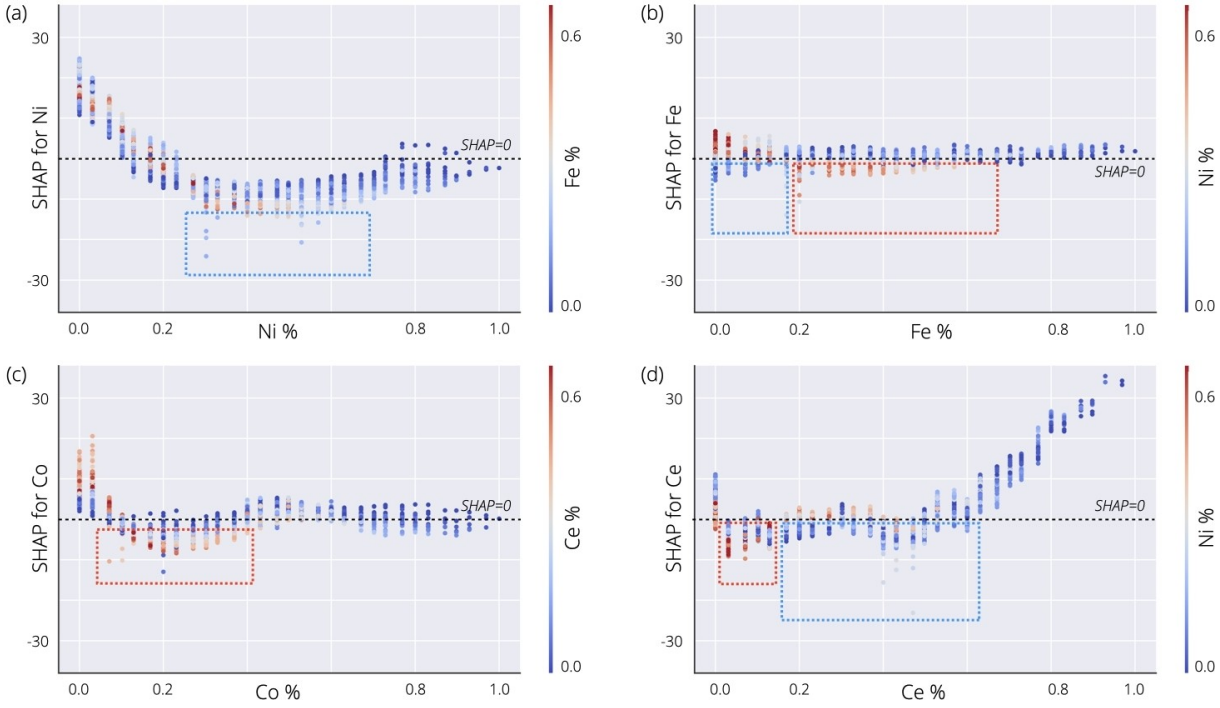

Figure 8. Feature dependencies of Ni, Fe, Co, and Ce. SHAP values are represented on the y-axis, while colors indicate the values of the corresponding interacting feature. For each feature, only the strongest interacting feature is shown: (a) Ni with Fe, (b) Fe with Ni, (c) Co with Ce, and (d) Ce with Ni. Regions with low OVP values are delineated and color–coded according to the values of interacting features. A dashed black line marks where SHAP value = 0. Dependency plots for all the other interactions are shown in Figure S8 to S11.

tion featuring Ni = 27.1 %, Fe = 7.1 %, Co = 49.8 %, and Ce = 16.9 % results in a predicted OVP of 416.12 mV. This prediction is primarily driven by Ni (with a negative impact), followed by Co (positive), Ce (negative), and Fe (negative) (Figure 9a). However, these effects vary across compositions, as shown in another example featuring Ni = 13.9 %, Fe = 23.1 %, Co = 3.1 %, and Ce = 60 %. In this case, the influence of each feature is relatively balanced, with Ni remaining the most influential (in a negative direction). Interestingly, despite having the largest proportion, Ce exerts the least influence in this composition (Figure 9b).

There are several alternative methods for generating explanations at both the local level (e.g., feature importance, learning process) and global level (e.g., feature interaction and impact). Our XAI framework for OER primarily focuses on providing post–hoc explanations using the SHAP algorithm. Alternative approaches include self–interpretable models (such as decision trees or logistic regression), single–decision explanations (such as Local Interpretable Model–Agnostic Explainer or LIME), and whole-model explanations (such as Concept Activation Vectors or CAVs).[21,23] Utilizing these methodologies would offer a wider and deeper perspective of the materials science

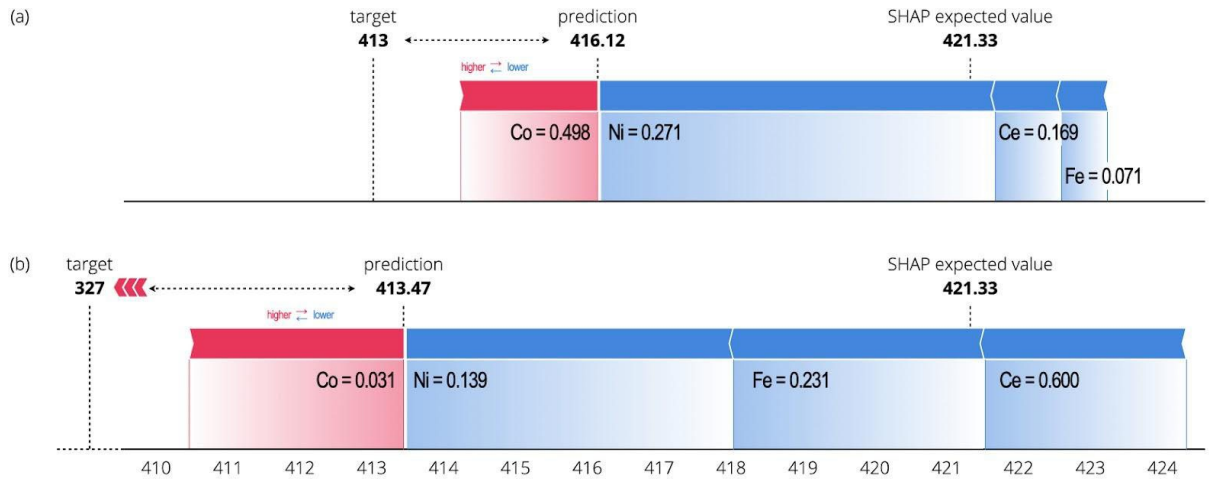

Figure 9. Additive force for random forest overpotential predictions on (a) normal data and (b) outlier.

problem and its AI-driven solution. Nevertheless, our results present a compelling case for further using XAI in materials science, particularly considering that AI and ML research in this domain is yet to reach its full potential.

Conclusions

Explainable artificial intelligence is the next significant step in catalyst design for oxygen evolution reactions. In summary, we introduce an OER XAI framework for predicting the overpotential of (Ni–Fe–Co–Ce)O_x catalysts. Our approach integrates a random forest model with Shapley additive explanation to build an accurate yet interpretable predictive model. The results indicate improvements in the predictive performance of existing machine learning models for (Ni–Fe–Co–Ce)O_x compositions, achieving an R^2 value of 0.8221. Moreover, the generated explanations regarding feature importance, interactions, and cumulative impact contribute to a more comprehensive understanding of how Ni, Fe, Co, and Ce influence the prediction of OVP. These insights will be valuable in optimizing catalyst synthesis processes, ultimately driving novel materials development and discovery. The link for our codes and data can be accessed in this link: www.bit.ly/xai-oer-2023.

Acknowledgements

Open Access funding provided by the Max Planck Society. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: catalyst · explainable AI · oxygen evolution reaction · random forest · SHAP

[1] C. M. Pelicano, M. Saruyama, R. Takahata, R. Sato, Y. Kitahama, H. Matsuzaki, T. Yamada, T. Hisatomi, K. Domen, T. Teranishi, *Adv. Funct. Mater.* **2022**, *32*, 2202987.
[2] T. Reier, H. N. Nong, D. Teschner, R. Schlogl, P. Strasser, *Adv. Energy Mater.* **2017**, *7*, 1601275.
[3] C. M. Pelicano, J. Li, M. Cabrero-Antonino, I. Silva, L. Peng, N. V. Tarakina, S. Navalon, H. Garcia, M. Antonietti, *J. Mater. Chem. A.* **2023**, DOI: 10.1039/D3TA05701A.
[4] C. M. Pelicano, H. Tong, *Appl. Res.* **2023**, e202300080. DOI: 10.1002/appl.202300080.
[5] M. Tahir, L. Pan, F. Idrees, X. W. Zhang, L. Wang, J. J. Zou, Z. L. Wang, *Nano Energy* **2017**, *37*, 136–157.
[6] L. Han, S. J. Dong, E. K. Wang, *Adv. Mater.* **2016**, *28*, 9266–9291.
[7] C. Liao, B. Yang, N. Zhang, M. Liu, G. Chen, X. Jiang, G. Chen, J. Yang, X. Liu, T. S. Chan, Y. J. Lu, R. Ma, W. Zhou, *Adv. Funct. Mater.* **2019**, *29*, 1904020.
[8] X. H. Xie, L. Du, L. T. Yon, S. Y. Park, Y. Qiu, J. Sokolowski, W. Wang, Y. Y. Shao, *Adv. Funct. Mater.* **2022**, *32*, 2110036.
[9] Y. Zhai, X. Ren, B. Wang, S. Liu, *Adv. Funct. Mater.* **2022**, *32*, 2207536.
[10] M. Chen, H. Li, C. Wu, Y. Liang, J. Qi, J. Li, E. Shangguan, W. Zhang, R. Cao, *Adv. Funct. Mater.* **2022**, *32*, 2206407.
[11] M. Saruyama, C. M. Pelicano, T. Teranishi, *Chem. Sci.* **2022**, *13*, 2824–2840.
[12] J. K. Nørskov, T. Bligaard, J. Rossmeisl, C. H. Christensen, *Nat. Chem.* **2009**, *1*, 37–46.
[13] G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa, A. Fazzio, *J. Phys. Mater.* **2019**, *2*, 032001.
[14] Z. Sun, H. Yin, K. Liu, S. Cheng, G. K. Li, S. Kawi, H. Zhao, G. Jia, Z. Yin, *SmartMat.* **2022**, *3*, 68–83.
[15] M. McBride, N. Persson, E. Reichmanis, M. Grover, *Processes* **2018**, *6*, 79.
[16] P. R. Regonia, C. M. Pelicano, R. Tani, A. Ishizumi, H. Yanagi, K. Ikeda, *Optik (Stuttg)* **2020**, *207*, 164469.
[17] P. R. Regonia, J. P. Olorocisimo, F. De Los Reyes, K. Ikeda, C. M. Pelicano, *NanoImpact* **2022**, *28*, 100442.
[18] G. H. Gu, J. Noh, I. Kim, Y. Jung, *J. Mater. Chem. A* **2019**, *7*, 17096.
[19] X. Jiang, Y. Wang, B. Jia, X. Qu, M. Qin, *ACS Appl. Mater. Interfaces* **2022**, *14*, 41141–41148.
[20] X. Mao, L. Wang, Y. Li, *J. Phys. Chem. Lett.* **2023**, *14*, 170–177.
[21] P. J. Phillips, C. A. Hahn, P. C. Fontana, A. Yates, K. Greene, D. Broniatowski, M. Przybock, *Natl. Inst. Stand. Tech. Interag Intern Rep.* **2021**, *8312*, 1–43.
[22] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, *Npj Comput Mater.* **2017**, *3*, 54.
[23] X. Zhong, B. Gallagher, S. Liu, B. Kailkhura, A. Hiszpanski, T. Y. J. Han, *Npj Comput Mater.* **2022**, *8*, 204.
[24] B. Esteki, M. Masoomi, M. Moosazadeh, C. Yoo, *Langmuir* **2023**, *39*, 4943–4958.
[25] J. A. Esterhuizen, B. R. Goldsmith, S. Linic, *Chem. Catal.* **2021**, *1*, 923–940.
[26] R. Palkovits, S. Palkovits, *ACS Catal.* **2019**, *9*, 8383–8387.
[27] X. Jiang, Y. Wang, B. Jia, X. Qu, M. Qin, *ACS Omega* **2022**, *7*, 14160–14164.
[28] S. Lundberg, S. I. Lee, *Adv. Neural Inf. Process* **2017**, *30*, 4765–4774.
[29] C. Chatfield, *Eur. J Oper. Res.* **1986**, *23*, 5–13.
[30] J. A. Haber, C. Xiang, D. Guevarra, S. Jung, J. Jin, J. M. Gregoire, *ChemElectroChem.* **2013**, *1*, 524–528.
[31] J. A. Haber, Y. Cai, S. Jung, C. Xiang, S. Mitrovic, J. Jin, A. T. Bell, J. M. Gregoire, *Energy Environ. Sci.* **2014**, *7*, 682–688.
[32] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, J. M. O'Sullivan, *Front. Bioinform.* **2022**, *2*, 927312.
[33] V. Barnett, T. Lewis, *Outliers in Statistical Data* John Wiley & Sons, **1994**.
[34] C. Isaksson, M. H. Dunham, *Proc. Int. Workshop Mach. Learn. Data Mining Pattern Recognit* **2009**, *5632*, 440–453.
[35] B. L. DeCost, J. R. Hattrick-Simpers, Z. Trautt, A. G. Kusne, E. M. Campo, M. A. Green, *Mach. Learn.: Sci. Technol.* **2020**, *1*, 033001.
[36] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, *559*, 547–555.
[37] A. Liaw, M. Wiener, *R News* **2002**, *2*, 18–22.
[38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
[39] G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R* 2nd ed. Springer, **2013**.
[40] R. G. Pontius, O. Thontteh, H. Chen, *Environ Ecol Stat.* **2007**, *15*, 111–142.
[41] K. Hirose, H. Masuda, *Entropy* **2018**, *20*, 632.
[42] S. Lundberg, S. I. Lee, *SHAP. GitHub.* Published August 17, **2023**. Accessed **2023**. https://github.com/shap/shap.
[43] A. J. Bard, L. R. Faulkner, *Electrochemical Methods: Fundamentals and Applications. 2nd ed.* Wiley Global Education, **2000**.
[44] J. A. Haber, E. Anzenburg, J. Yano, C. Kisielowski, J. M. Gregoire, *Adv. Energy Mater.* **2015**, *5*, 1402307.
[45] L. Wang, J. Geng, W. Wang, C. Yuan, L. Kuai, B. Geng, *Nano Res.* **2015**, *8*, 3815–3822.
[46] D. A. Corrigan, *J. Electrochem. Soc.* **1987**, *134*, 377.
[47] D. A. Corrigan, R. M. Bendert, *J. Electrochem. Soc.* **1989**, *136*, 723.
[48] C. C. L. McCrory, S. H. Jung, J. C. Peters, T. F. Jaramillo, *J. Am. Chem. Soc.* **2013**, *135*, 16977.

[49] R. D. L. Smith, M. S. Prévot, R. D. Fagan, Z. Zhang, P. A. Sedach, M. K. J. Siu, S. Trudel, C. P. Berlinguette, *Science* **2013**, *340*, 60.
[50] D. M. F. Santos, L. Amaral, B. Šljukić, D. Macciò, A. Saccone, C. A. C. Sequeira, *J. Electrochem. Soc.* **2014**, *161*, F386.
[51] M. W. Louie, A. T. Bell, *J. Am. Chem. Soc.* **2013**, *135*, 12329.
[52] F. Song, M. M. Busch, B. Lassalle-Kaiser, C. S. Hsu, E. Petkucheva, M. Bensimon, H. M. Chen, C. Corminboeuf, X. Hu, *ACS Cent. Sci.* **2019**, *5*, 558–568.
[53] L. Trotochaud, S. L. Young, J. K. Ranney, S. W. Boettcher, *J. Am. Chem. Soc.* **2014**, *136*, 6744–6753.
[54] D. Friebel, M. W. Louie, M. Bajdich, K. E. Sanwald, Y. Cai, A. M. Wise, M.-J. Cheng, D. Sokaras, T.-C. Weng, R. Alonso-Mori, R. C. Davis, J. R. Bargar, J. K. Nørskov, A. Nilsson, A. T. Bell, *J. Am. Chem. Soc.* **2015**, *137*, 1306.

Manuscript received: November 30, 2023
Revised manuscript received: February 23, 2024
Version of record online: March 8, 2024