Prediction of Oxygen Evolution Activity for NiCoFe Oxide Catalysts via Machine Learning

Xue Jiang,[§] Yong Wang,[§] Baorui Jia,* Xuanhui Qu, and Mingli Qin*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Transition metal (such as Fe, Co, and Ni) oxides are excellent systems in the oxygen evolution reaction (OER) for the development of non-noble-metal-based catalysts. However, direct experimental evidence and the physical mechanism of a quantitative relationship between physical factors and oxygen evolution activity are still lacking, which makes it difficult to theoretically and accurately predict the oxygen evolution activity. In this work, a data-driven method for the prediction of overpotential (OP) for (Ni-Fe-Co)O_x catalysts is proposed via machine learning. The physical features that are more related to the OP for the OER have been constructed and analyzed. The random forest regression model works exceedingly well on OP prediction with a mean relative error of 1.20%. The features based on first ionization energies (FIEs) and outermost d-orbital electron numbers (DEs) are the principal factors and their variances (δ_{FIE} and δ_{DE}) exhibit a linearly decreasing correlation with OP, which gives direct guidance for an OP-oriented component design. This method provides novel and promising insights for the prediction of oxygen evolution activity and physical factor analysis in (Ni-Fe-Co)O_x catalysts.

Ni-Fe-Co oxide oxygen evolution catalyst
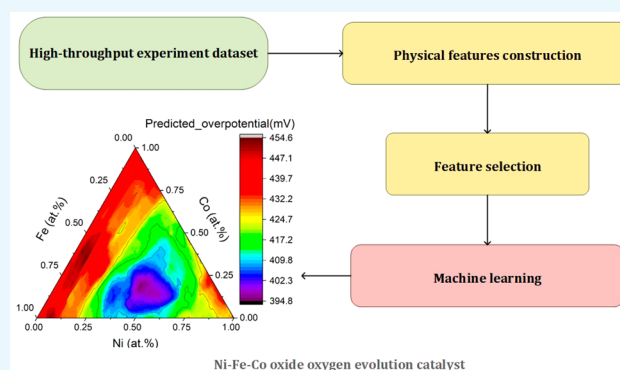
Catalysts play a critical role in oxygen electrochemical processes for renewable energy storage and conversion devices such as fuel cells, artificial photosynthesis, and metal−air batteries.[1−6] For example, the oxygen evolution reaction (OER) involves a four-electron (4e⁻) transfer, and such a complicated process results in sluggish reaction kinetics.[7−9] Therefore, the OER is considered as the main bottleneck toward the practical implementation of polymer electrolyte membrane (PEM) electrolysis and water splitting. Noble-metal (Ru and Ir) catalysts exhibit high OER activity and can reduce the required overpotential; however, their scarcity and high cost limit their wide application.[10−12] Transition metal (such as Fe, Co, and Ni) oxides are excellent systems in the OER for the development of non-noble-metal-based catalysts.[13] The quantitative prediction of OER activity is critical for transition metal oxide catalyst design.[14] Due to fact that direct experimental evidence and the mechanism of the quantitative relationship between physical factors and oxygen evolution activity from physical insights are not yet clear, OER activity prediction remains an unsolved challenge.[13] Usually, scientists calculate the adsorption energetics of different chemical structures by density functional theory and indirectly infer the OER activity, which is more suitable for rationalizing observed activity trends and facts rather than predicting them in advance from the large potential chemical space. Consequently, a predictive method as the function of latent influencing factors needs to be captured for OER activity.

Machine-learning (ML) approaches are transforming materials research by changing the paradigm from "trial and error" to a data-driven methodology, thereby accelerating the discovery of new materials.[15−25] Recently, the catalyst community has begun to utilize ML tools to accelerate the overpotential prediction of the oxygen evolution reaction for single atoms,[26] forecast Ni-Co-Fe-Ce water oxidation catalysts[20] and evaluate the perovskite chemistry factors of OER activity.[14] Various factors of oxide perovskite catalysts have been demonstrated over the past 60 years, such as the reaction free energy and e_g occupancy, which were obtained by DFT calculations. A simple factor, μ/t, was derived from symbolic regression for perovskite catalysts.[27] A good factor should be simple and yet provide physical insight to therefore guide and accelerate the discovery of new oxide OER catalysts. Despite the great potential, its use has been notably absent in transition metal oxide systems, such as (Ni-Fe-Co)O_x catalysts.

In this work, we introduce a data-driven method to predict the overpotential (OP) for (Ni-Fe-Co)O_x catalysts using a

machine-learning algorithm. The relationship between multiple physical features and OP properties was successfully determined, by considering valence electron number, relative atomic mass, atomic number, atomic radius (nonbonded), covalent radius, ionization energies (first), electron affinity, electronegativity (Pauling scale), and outermost d-orbital electron number. The random forest regression model works exceedingly well with a mean relative error of 1.20% evaluated on a hold-out set. The importance of physical features has been further analyzed. The first ionization energies (FIEs) and outermost d-orbital electron numbers (DE) are the principal factors, and their variances (δ_{FIE} and δ_{DE}) exhibit a linearly decreasing correlation with OP. They give a direct guidance for an OP-oriented component design for (Ni-Fe-Co)O$_x$ catalysts. Our work aims to provide novel and promising physical insight for OER activity of (Ni-Fe-Co)O$_x$ catalysts.

All of the data used in this work were collected and screened from the published studies of Haber by a high-throughput experiment.[27] We consider the oxide catalysts that belong to the Ni$_x$Co$_y$Fe$_z$ system, where the mole fractions of each element of x, y, and z is constrained by $x + y + z = 100\%$. The data set consists of 496 entries (see the Supporting Information for details), covering the elemental composition in percent representing different (Ni-Fe-Co)O$_x$ materials and characterizing their overpotential (OPs) using a 10 s chronopotentiometry measurement at 10 mA/cm^2 in O$_2$-saturated 1.0 M NaOH(aq).[27] The overall data set therefore possesses three features (input variables) represented by composition with respect to the elements and one target OP (output variable). Figure 1 gives the whole data visualization
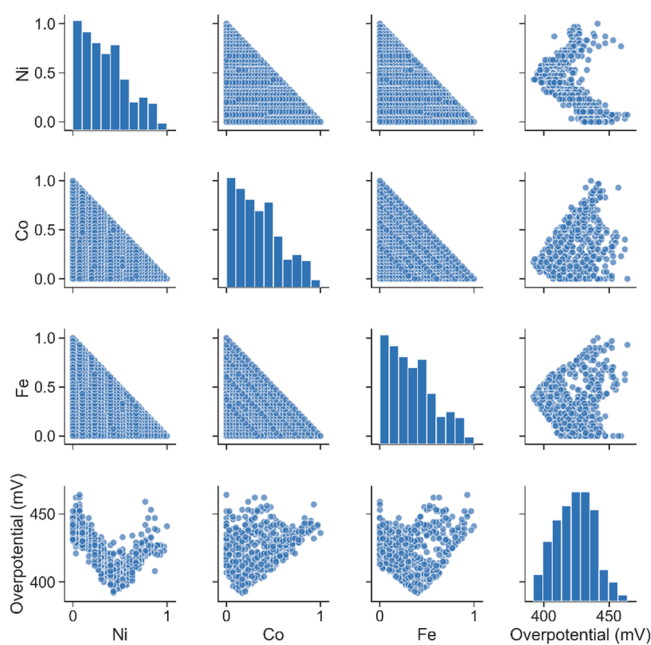
Figure 1. Original data set visualization by categorical plots of Ni, Co, Fe, and overpotential.

carried out with the Python programming language[28] and the statistical data visualization library Seaborn.[29] The composition for each of Ni, Co, and Fe ranges from 0 to 1 by even steps of 3.33 atom %, and covers the whole composition space that may be formed. As the composition for each of Ni, Co, and Fe increases, the OP shows an overall trend of first decreasing and then increasing. This method can obviously show the optimal

composition combination that minimizes the OPs of metal oxide catalysts. Complete and comprehensive data provide a reliable basis for a prediction model and physical fact analysis.

Physical features are critical for representing the intrinsic relationship between the latent fact and OP property. From the perspective of the atom level, nine accessible primary physical features by empirical experience given in Table 1 are used, including the valence electron number, relative atomic mass, atomic number, nonbonded atomic radius, covalent radius, first ionization energy, electron affinity, Pauling scale electronegativity, and outermost d-orbital electron number. For Ni, Co, and Fe elements, their physical properties corresponding with these features are collected from The Royal Society of Chemistry's interactive periodic table database.[30] The composition (C_i) and the associated elemental properties were adopted to numerically represent each catalyst sample with the featured transformation functions of eqs 1 and 2 for the purpose of converting the original chemical element space to a primary physical features space. For each catalyst sample in the 496 entries, $\overline{X}$ calculates the weighted average of the element content corresponding to each physical feature, and δ_X produces the variance for each physical feature reflecting the physical difference of chemical element. In eqs 1 and 2, C_i is the mole fraction of each element and P_i corresponds to the properties of each element, respectively. Thus, 18 descriptive features that may be physically relevant to the OP are constructed by $\overline{X}$ and δ_X. The original data set is transformed into a new data set with 496 × 18 shape (see the Supporting Information for details).

In order to remove the linear correlation between two variables, Pearson correlation coefficients were then calculated before machine learning. Pearson correlation coefficients can be described as

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}$$

(3)

where R is the correlation coefficient matrix, C is the covariance matrix. Covariance indicates the level to which two variables vary together, and it belongs to [-1,1]. The closer the $|R_{ij}|$ value is to 1, the higher the linear correlation between the two variables is. The heatmap of Pearson correlation coefficients for the physical features and OP are shown in Figure 2. Features with an absolute value of correlation coefficient greater than 0.95 were considered highly correlated (the boxes in dark blue and dark red in Figure 2). In the highly correlated feature pairs, one can be linearly expressed and replaced by the other. For example, the correlation coefficient between $\overline{VEN}$ and $\overline{EA}$ is 1; thus, $\overline{EA}$ is retained instead of $\overline{VEN}$. Therefore, $\overline{VEN}$, $\overline{AN}$, $\overline{EA}$, δ_{VEN}, δ_{RAM}, δ_{AN}, δ_{RA}, and δ_{EA} are excluded. Finally, the data set contains 496 entries with 10 features ($\overline{RAM}$, $\overline{RA}$, $\overline{RC}$, $\overline{FIE}$, $\overline{EP}$, $\overline{DE}$, δ_{RC}, δ_{FIE}, δ_{EP} and δ_{DE}) and 1 target property.

The OP is plotted as a function of $\overline{RAM}$, $\overline{RA}$, $\overline{RC}$, $\overline{FIE}$, $\overline{EP}$, $\overline{DE}$, δ_{RC}, δ_{FIE}, δ_{EP} and δ_{DE}, respectively, as shown in Figure 3. It is obvious that OP decreases first and then rises with the increase of $\overline{FIE}$ and $\overline{DE}$. The optimal values of $\overline{FIE}$ and $\overline{DE}$ may be set at around 750 and 7 to achieve the lowest OP. As δ_{RC}, δ_{FIE}, δ_{EP}, and δ_{DE} increase, OP decreases, and the phenomenon is more obvious for δ_{FIE} and δ_{DE}.

Machine-learning algorithms can fit a data-driven over-potential model with the selected physical features as the input and the target property OP as the output. Before model

Table 1. Material Physical Features, Abbreviations, Units and the Transformation Formulas

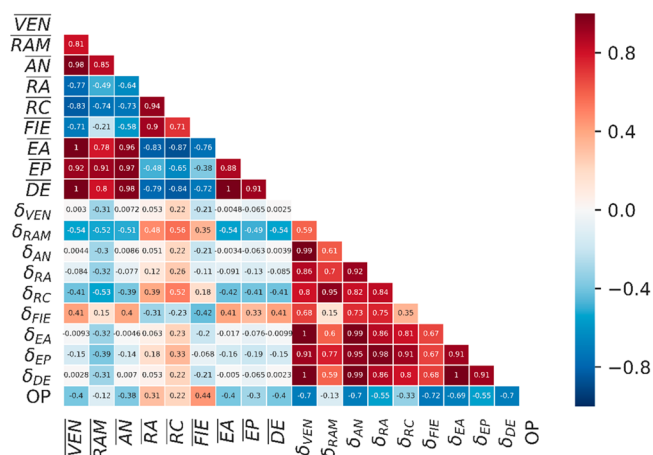| features	abbreviation	unit	formula
valence electron number	VEN		
relative atomic mass	RAM		
atomic number	AN		
atomic radius, nonbonded	RA	Å	$\bar{X} = \sum\limits_{i \in \{Ni, Co, Fe\}} C_i P_i$ (1)
covalent radius	RC	Å	
first Ionization energy	FIE	kJ mol^{-1}	$\delta_X = \sum\limits_{i \in \{Ni, Co, Fe\}} [P_i - \bar{X}]^2 C_i$ (2)
electron affinity	EA	kJ mol^{-1}	
electronegativity (Pauling scale)	EP	-	
outermost d-orbital electron number	DE	-	

Figure 2. Heat map of Pearson correlation coefficients for the physical features and OP.

construction, model selection and parameter optimization ensure which algorithm together with the suitable parameters would perform the best and be considered. The data set with 10 features and 1 target property was split by 80% (397 data entries) as the training set and 20% (99 data entries) left as the testing set (hold-out set). Several well-known machine-learning algorithms were used, such as stochastic gradient descent regression (SGDR) with penalties of L1 and L2, lasso regression (Lasso), elastic net regression (ElasticNet), multi-

layer proceptron (MLP), tree regression (TreeR), Adaboost regression (AdaBR), gradient boosting regression (GBR), random forest regression (RFR), logistic regression (LogisticR), kernel ridge regression (KernelRidge), Bayesian Ridge regression (BayesRidge), support vector regression (SVR) with radial basis function kernel, and k-nearest neighbor regression (KNR). Parameter tuning was performed by a grid search on the training set by 5-fold cross validation for each machine-learning model, and the parameter with the best average mean squared error was determined. Then, the model was trained with the best parameter on the training set. Figure 4 shows the mean squared errors (MSE) and the standard deviations for different models during model selection. GBR, RFR, KernelRidge and KNR models exhibit excellent MSEs and uncertainties, and among them RFR has the lowest MSE of 40.6.

Then, we retrained the RFR model with the optimized parameters on the training set and evaluated the metrics of MSE and mean relative error (MRE) on 20% testing set, respectively. Figure 5a shows the diagonal scatter plot for the predicted OP and the ground truth by the RFR model during training and testing on the basis of the transformed and selected data set. RFR performs the best with an MRE value of 1.20% and an MSE value of 49.79. Figure 5b shows the contour map of the predicted overpotential by the RFR model under different compositions. The RFR model performs well on the "unseen" hold-out data set, whose MSE value is only 9.19 higher than that in training process, illustrating that this
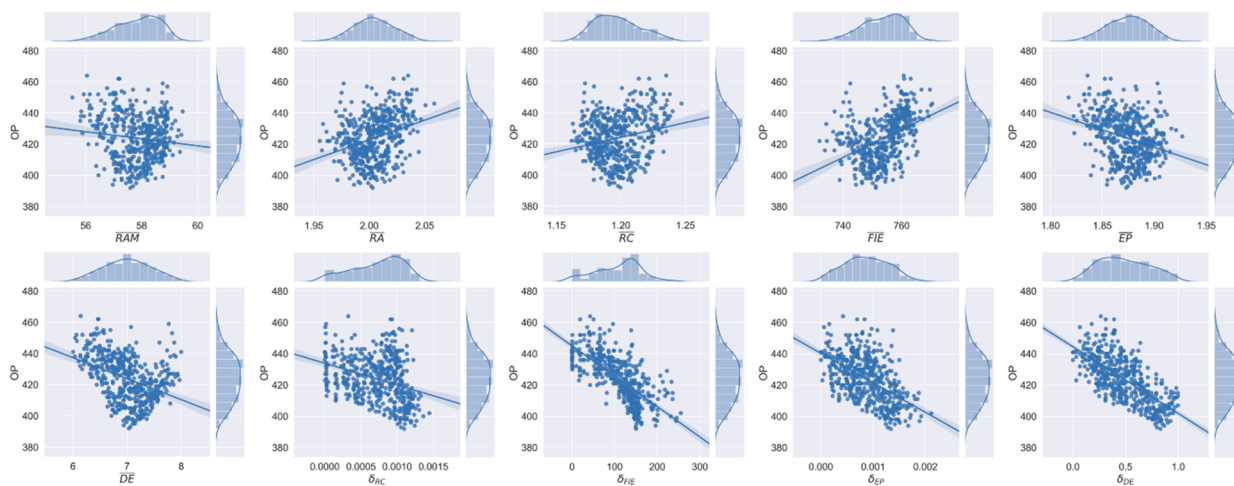
Figure 3. OP as a function of $\overline{RAM}$, $\overline{RA}$, $\overline{RC}$, $\overline{FIE}$, $\overline{EP}$, $\overline{DE}$, δ_{RC}, δ_{FIE}, δ_{EP}, and δ_{DE}.
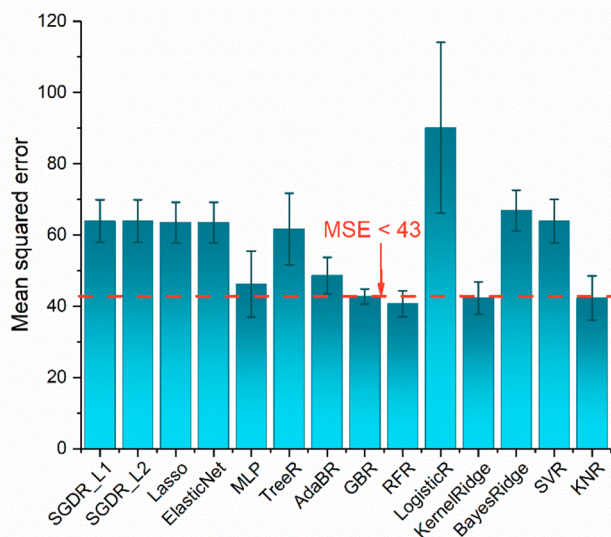
Figure 4. Mean squared errors for different models during model selection.

model can be used for OP prediction with generalization capability in (Ni-Fe-Co)O$_x$ catalysts.

Next the contributions each physical feature makes to the high-precision OP prediction model will be revealed. On the basis of the trained RFR model, the 10 physical features are ranked by their feature importance in Figure 5c. The greater contribution to high-precision OP prediction one feature makes, the higher the importance index is. δ_{FIE} and δ_{DE} are the most critical factors supported by the RFR model, and from the perspective of model precision, δ_{FIE} is more important than δ_{DE}. The importance indices of $\overline{\text{DE}}$ and $\overline{\text{FIE}}$ are relatively lower than that of δ_{FIE} and δ_{DE} but higher than the others. When the trends of the physical features (δ_{FIE}, δ_{DE}, $\overline{\text{DE}}$ and $\overline{\text{FIE}}$) in Figures 3 and 5 are combined, the δ_{FIE} and δ_{DE} values can be used to give direct guidance to design an optimal component for Ni, Co, and Fe for excellent OP, by adjusting the contents of different elements to make the two simple factors δ_{FIE} and δ_{DE} greater.

In summary, a data-driven approach to predict the OP for (Ni-Fe-Co)O$_x$ catalysts is proposed via machine learning. The physical features that are more related to the catalyst overpotential for the OER are constructed, covering valence electron number, relative atomic mass, atomic number, atomic radius (nonbonded), covalent radius, first ionization energy, electron affinity, electronegativity (Pauling scale), and outermost d-orbital electron number. The random forest regression

model works exceedingly well with a mean relative error of 1.20%. The simple and easily accessed factors (δ_{FIE} and δ_{DE}) by the variances of first ionization energies (FIE) and outermost d-orbital electron number (DE) importance are captured, exhibiting a linearly decreasing correlation with OP. They give direct guidance for the OP-oriented component design for (Ni-Fe-Co)O$_x$ catalysts. Our work aims to provide novel and promising physical insights into the OER activity of (Ni-Fe-Co)O$_x$ catalysts.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c00776.

Original data used in this work and the transformed data set with physical features (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Baorui Jia − *Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, People's Republic of China;* ⓘ orcid.org/0000-0002-1846-6971; Email: jiabaorui@ustb.edu.cn

Mingli Qin − *Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, People's Republic of China;* ⓘ orcid.org/0000-0002-4001-6539; Email: qinml@mater.ustb.edu.cn

Authors

Xue Jiang − *Beijing Advanced Innovation Center for Materials Genome Engineering, Collaborative Innovation Center of Steel Technology, University of Science and Technology Beijing, Beijing 100083, People's Republic of China*

Yong Wang − *Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, People's Republic of China*

Xuanhui Qu − *Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, People's Republic of China*
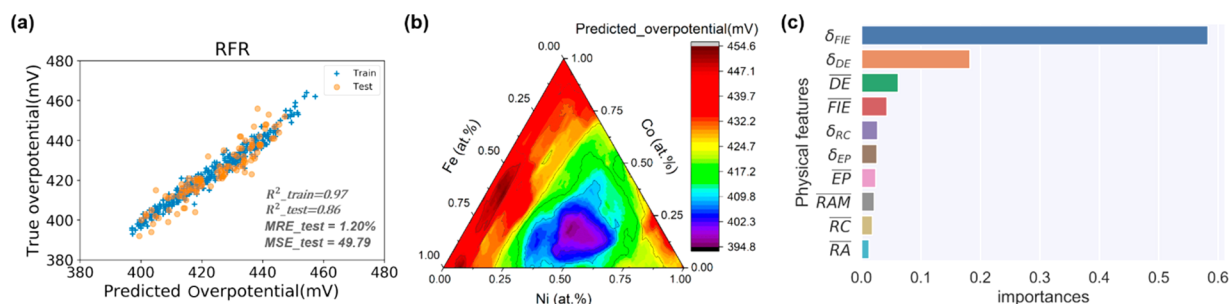
Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.2c00776

Author Contributions

§X.J. and Y.W. contributed equally to this work.

Notes

The authors declare no competing financial interest.

Figure 5. Machine-learning model by RFR. (a) Diagonal scatter plot for the predicted OP and the ground truth by RFR. (b) Contour map of the predicted overpotential by the RFR model under different compositions. (c) Physical feature importance ranking by the RFR model.

■ REFERENCES

(1) Wang, H. F.; Chen, L.; Pang, H.; Kaskel, S.; Xu, Q. MOF-derived electrocatalysts for oxygen reduction, oxygen evolution and hydrogen evolution reactions. *Chem. Soc. Rev.* **2020**, *49*, 1414−1448.

(2) Shang, H.; Sun, W.; Sui, R.; Pei, J.; Zheng, L.; Dong, J.; Jiang, Z.; Zhou, D.; Zhuang, Z.; Chen, W.; et al. Engineering Isolated Mn−N$_2$C$_2$ Atomic Interface Sites for Efficient Bifunctional Oxygen Reduction and Evolution Reaction. *Nano Lett.* **2020**, *20*, 5443−5450.

(3) Vij, V.; Sultan, S.; Harzandi, A. M.; Meena, A.; Tiwari, J. N.; Lee, W.-G.; Yoon, T.; KS, K. Nickel-based electrocatalysts for energy-related applications: oxygen reduction, oxygen evolution, and hydrogen evolution reactions. *ACS Catal.* **2017**, *7*, 7196−7225.

(4) Gao, R.; Dai, Q.; Du, F.; Yan, D.; Dai, L. C_(60)-Adsorbed Single-Walled Carbon Nanotubes as Metal-Free, pH-Universal, and Multifunctional Catalysts for Oxygen Reduction, Oxygen Evolution, and Hydrogen Evolution. *J. Am. Chem. Soc.* **2019**, *141*, 11658−11666.

(5) Han, X.; Ling, X.; Yu, D.; Xie, D.; Li, L.; Peng, S.; Zhong, C.; Zhao, N.; Deng, Y.; Hu, W. Atomically Dispersed Binary Co-Ni Sites in Nitrogen-Doped Hollow Carbon Nanocubes for Reversible Oxygen Reduction and Evolution. *Adv. Mater.* **2019**, *31*, 1905622.

(6) Surendran, S.; Shanmugapriya, S.; Sivanantham, A.; Shanmugam, S.; Selvan, R. K. Electrospun Carbon Nanofibers Encapsulated with NiCoP: A Multifunctional Electrode for Supercapattery and Oxygen Reduction, Oxygen Evolution, and Hydrogen Evolution Reactions. *Adv. Energy Mater.* **2018**, *8*, 1800555.

(7) Tran-Phu, T.; Daiyan, R.; Leverett, J.; Fusco, Z.; Tadich, A.; Di Bernardo, I.; Kiy, A.; Truong, T. N.; Zhang, Q.; Chen, H.; et al. Understanding the activity and stability of flame-made Co3O4 spinels: A route towards the scalable production of highly performing OER electrocatalysts. *Chem. Eng. J.* **2022**, *429*, 132180.

(8) Ali Khan, M. H.; Daiyan, R.; Han, Z.; Hablutzel, M.; Haque, N.; Amal, R.; MacGill, I. Designing Optimal Integrated Electricity Supply Configurations for Renewable Hydrogen Generation in Australia. *IScience* **2021**, *24*, 102539.

(9) Yates, J.; Daiyan, R.; Patterson, R.; Egan, R.; Amal, R.; Ho-Baille, A.; Chang, N. L. Techno-economic Analysis of Hydrogen Electrolysis from Off-Grid Stand-Alone Photovoltaics Incorporating Uncertainty Analysis. *Cell Reports Physical Science* **2020**, *1*, 100209.

(10) Qin, M.; Li, S.; Zhao, Y.; Lao, C. Y.; Qu, X.; et al. Unprecedented Synthesis of Holey 2D Layered Double Hydroxide Nanomesh for Enhanced Oxygen Evolution. *Adv. Energy Mater.* **2019**, *9*, 1803060.

(11) Zhang, Q.; Kumar, P.; Zhu, X.; Daiyan, R.; Bedford, N. M.; Wu, K.-H.; Han, Z.; Zhang, T.; Amal, R.; Lu, X. Electronically Modified Atomic Sites Within a Multicomponent Co/Cu Composite for Efficient Oxygen Electroreduction. *Adv. Energy Mater.* **2021**, *11*, 2100303.

(12) Daiyan, R.; Macgill, I.; Amal, R. Opportunities and Challenges for Renewable Power-to-X. *ACS Energy Letters* **2020**, *5*, 3843−3847.

(13) Hong, W. T.; Risch, M.; Stoerzinger, K. A.; Grimaud, A.; Suntivich, J.; Shao-Horn, Y. Toward the rational design of non-precious transition metal oxides for oxygen electrocatalysis. *Energy Environ. Sci.* **2015**, *8*, 1404−1427.

(14) Hong, W. T.; Welsch, R. E.; Shao-Horn, Y. Descriptors of Oxygen-Evolution Activity for Oxides: A Statistical Evaluation. *J. Phys. Chem. C* **2016**, *120*, 78−86.

(15) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Norquist, A. J.; et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* **2016**, *533*, 73−76.

(16) Xue, D.; Balachandran, P. V.; Hogden, J.; Theiler, J.; Xue, D.; Lookman, T. Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* **2016**, *7*, 11241.

(17) Xue, D.; Xue, D.; Yuan, R.; Zhou, Y.; Balachandran, P. V.; Ding, X.; Sun, J.; Lookman, T. An informatics approach to transformation temperatures of NiTi-based shape memory alloys. *Acta Mater.* **2017**, *125*, 532−541.

(18) Jiang, X.; Yin, H. Q.; Zhang, C.; Zhang, R. J.; Zhang, K. Q.; Deng, Z. H.; Liu, G. Q.; Qu, X. H. An materials informatics approach to Ni-based single crystal superalloys lattice misfit prediction. *Comput. Mater. Sci.* **2018**, *143*, 295−300.

(19) Yuan, R.; Liu, Z.; Balachandran, P. V.; Xue, D.; Zhou, Y.; Ding, X.; Sun, J.; Xue, D.; Lookman, T. Accelerated Discovery of Large Electrostrains in BaTiO3-Based Piezoelectrics Using Active Learning. *Adv. Mater.* **2018**, *30*, 1702884.

(20) Palkovits, R.; Palkovits, S. Using Artificial Intelligence To Forecast Water Oxidation Catalysts. *ACS Catal.* **2019**, *9*, 8383−8387.

(21) Tshitoyan, V.; Dagdelen, J.; Weston, L.; Dunn, A.; Rong, Z.; Kononova, O.; Persson, K. A.; Ceder, G.; Jain, A. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **2019**, *571*, 95−98.

(22) Jiang, X.; Jia, B.; Zhang, G.; Zhang, C.; Ma, H. A strategy combining machine learning and multiscale calculation to predict tensile strength for pearlitic steel wires with industrial data. *Scripta Mater.* **2020**, *186*, 272−277.

(23) Zhang, Y.; Wen, C.; Wang, C.; Antonov, S.; Su, Y.; et al. Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. *Acta Mater.* **2020**, *185*, 528−539.

(24) Li, S.; Liu, Y.; Chen, D.; Jiang, Y.; Nie, Z.; Pan, F. Encoding the atomic structure for machine learning in materials science. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2022**, *12*, e1558.

(25) Wen, C.; Wang, C.; Zhang, Y.; Antonov, S.; Su, Y.; et al. Modeling solid solution strengthening in high entropy alloys using machine learning. *Acta Mater.* **2021**, *212*, 116917.

(26) Wu, L.; Guo, T.; Li, T. Machine learning accelerated prediction of overpotential of oxygen evolution reaction of single atom catalysts. *iScience* **2021**, *24*, 102398.

(27) Haber, J. A.; Cai, Y.; Jung, S.; Xiang, C.; Mitrovic, S.; Jin, J.; Bell, A. T.; Gregoire, J. M. Discovering Ce-rich oxygen evolution catalysts, from high throughput screening to water electrolysis. *Energy Environ. Sci.* **2014**, *7*, 682−688.

(28) Oliphant, T. E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 10−20.

(29) Waskom, M. Seaborn: statistical data visualization. *Journal of Open Source Software* **2021**, *6*, 3021.

(30) The Royal Society of Chemistry's interactive periodic table database (available from: https://www.rsc.org/periodic-table/).