Journal of Materials Chemistry A

# Directly Predicting Limiting Potentials from Easily Obtainable Physical Properties of Graphene-Supported Single-Atom Electrocatalysts by Machine Learning

| | |
|---|---|
| Journal: | *Journal of Materials Chemistry A* |
| Manuscript ID | TA-ART-12-2019-013404.R1 |
| Article Type: | Paper |
| Date Submitted by the Author: | 20-Jan-2020 |
| Complete List of Authors: | Lin, Shiru; University of Puerto Rico, Rio Piedras Campus,<br>Xu, Haoxiang; Beijing University of Chemical Technology, State Key Laboratory of Organic-Inorganic Composites<br>Wang, Yekun; University of Puerto Rico Rio Piedras<br>Zeng, Xiao Cheng; University of Nebraska-Lincoln, Department of Chemistry<br>Chen, Zhongfang; University of Puerto Rico, Department of Chemistry |
| | |

# Directly Predicting Limiting Potentials from Easily Obtainable Physical Properties of Graphene-Supported Single-Atom Electrocatalysts by Machine Learning

Shiru Lin[†], Haoxiang Xu[‡,ǁ], Yekun Wang[§], Xiao Cheng Zeng,[ǁ] Zhongfang Chen[†]*

[†] Department of Chemistry, University of Puerto Rico, San Juan, PR 00931, USA

[‡] Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing, China

[§] Department of Mathematics, University of Puerto Rico, Rio Piedras, San Juan, PR 00931, USA

[ǁ] Department of Chemistry, Department of Chemical & Biomolecular Engineering, and Department of Mechanical & Materials Engineering, University of Nebraska, Lincoln, NE, USA

*Corresponding author E-mail: zhongfangchen@gmail.com (ZC)

**ABSTRACT**

Oxygen reduction reaction (ORR), oxygen evolution reaction (OER), and hydrogen evolution reaction (HER) are three critical reactions for energy-related applications, such as water electrolyzers and metal-air batteries. Graphene-supported single-atom catalysts (SACs) have been widely explored; however, either experiments or density functional theory (DFT) computations cannot screen catalysts at high speed. Herein, based on DFT computations of 104 graphene-supported SACs ($M@C_3$, $M@C_4$, M@pyridine-$N_4$, and M@pyrrole-$N_4$), we built up machine learning (ML) models to describe the underlying pattern of easily obtainable physical properties and limiting potentials (mean square errors = 0.027/0.021/0.035 V for ORR/OER/HER, respectively), and employed these models to predict the catalytic performance of 260 other graphene-supported SACs containing metal-$N_xC_y$ active sites ($M@N_xC_y$). We recomputed the top catalysts recommended by ML towards ORR/OER/HER by DFT, which confirmed the reliability of our ML model, and identified two OER catalysts (Ir@pyridine-$N_3C_1$ and Ir@pyridine-$N_2C_2$) outperforming noble metal oxides, $RuO_2$ and $IrO_2$. The ML models quantitatively unveiled the significance of various descriptors and fast narrowed down the candidate list of graphene-supported single-atom catalysts. This approach can be easily used to screen and design other SACs, and significantly accelerate the catalyst design for many other important reactions.

**KEYWORDS:** Single-Atom Catalysts, Machine Learning, Limiting Potential, ORR/OER/HER, Density Functional Theory (DFT).

## 1. INTRODUCTION

Oxygen reduction reaction (ORR) oxygen evolution reaction (OER), and hydrogen evolution reaction (HER) are among the core electrochemical processes in clean energy conversion and storage devices, such as metal-air batteries, [1,2] water electrolysers,[3] and fuel cells.[4-6] However, their inherent reaction rates are rather sluggish. Noble metal (Pd and Pt) and noble metal oxide ($IrO_2$) are state-of-the-art electrocatalysts, [7,8] but the high cost and scarcity of noble metal materials limit their large-scale and sustainable applications. Thus, developing stable, low-cost, and high-performance catalysts for these reactions are highly desirable.

Single-atom catalysts (SACs), in which the well-dispersed isolated metal atoms are anchored on appropriate substrates, have emerged as a new frontier of heterogeneous catalysts due to their highly increased coverage of active sites, much enhanced catalytic performance, and maximal (100%) metal utilization.[9-15] Especially, metal-nitrogen-carbon (M-N-C) SACs, where a transition metal atom (M = Co, Fe, Ni, Mn, etc.) is located at the center of nitrogen (N) doped graphene support (C), showed great promise as substitutes for precious metal electrocatalysts.[16-23] At least hundreds of M-N-C SACs exist due to a sequence of physical structural variables, such as different transition metals and various N/C combinations. Therefore, trial-and-error approaches are rather inadequate to search for highly efficient catalysts in a reasonable time scale. Though volcano curves and approximate linear relationships[24,25] between some single factors and the performances of catalysts were

put forward, the rather simple linear relationships can only give rather rough judgments, instead of direct predictions of their catalytic performance.

Machine learning (ML) is an efficient statistic method, which builds up models based on input data[26] and computer algorithms to output the desired information.[27-30] Machine learning can be employed to depict the intricate relations between descriptors and performance, analyze the importance of each descriptor, and predict the performance of a large number of unknown systems.

The strong ability of ML has been applied to fast screening of materials with specific properties[31-40] and catalysts with high performance.[41-47] So far, ML in catalysis is mainly focused on two aspects: (1) establishing the correlations of physical properties and adsorption strength of reaction intermediates (Figure 1, Path I), because the catalytic activity is significantly affected by the binding strength. For instance, by the least absolute shrinkage and selection operator (LASSO) regressions,[48] O'Connor *et al.* revealed that the interfacial binding strength of single metal atom on oxide supports is correlated with readily available physical properties of both the absorbed metal (such as oxophilicity) and the support (such as reducibility);[41] (2) Identifying the relationships between the intermediate adsorption strengths and the performance of the catalyst (Figure 1, Path II). For instance, Ma *et al.* demonstrated that ML can well estimate the CO adsorption energy on the surface of multimetallic alloys, which can be used as a descriptor to screen $CO_2$ electroreduction catalysts.[49] However, the physical properties are intrinsic and

easy-obtained, and the limiting potential ($U_L$) is the most direct standard for the activity of an electrocatalyst, then, is it possible for us to directly predict the limiting potentials based on the physical properties (Figure 1, Path III)? The powerful ML algorithms may help find correlations between physical properties and limiting potentials. If successful, the discovery of electrocatalysts can be remarkably accelerated.

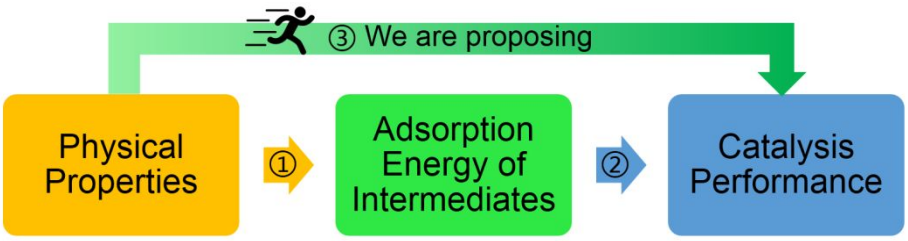



**Figure 1.** The schematic of three different aspects for catalyst development using ML techniques

Herein, by taking advantage of the ML algorithm and DFT computations, we depicted the underlying pattern of the physical properties of 104 graphene-supported SACs and their limiting potentials towards ORR/OER/HER reactions. The ML models for these three reactions were further used to predict the catalytic performance of 260 other graphene-supported metal-nitrogen/carbon systems ($M@N_xC_y$). The reliability of the ML models was confirmed by the DFT computed limiting potential ($U_L$) values of the top ML-recommended electrocatalysts (0.61, 1.51, 0.003 V for ORR, OER, and HER). We further quantitatively unveiled the significance of various

descriptors for M-N-C catalysts, which can significantly narrow down the candidate lists of $M@N_xC_y$ SACs and provide deep insights and guidance towards future catalyst design. This work provides a new paradigm for directly predicting the catalytic performance from physical properties of catalyst candidates, and vividly demonstrates the strong capability of ML in screening and design of catalysts.

## 2. COMPUTATIONAL METHODS

The training data of graphene-supported SACs were obtained from the previous study,[21] which include 104 graphene-supported structure compositions ($M@C_3$, $M@C_4$, M@pyridine-$N_4$, and M@pyrrole-$N_4$) and their $U_L$ values for ORR, OER and HER of each system.

Random Forest (RF),[50,51] a widely used integrated algorithm of Decision Tree[52] as implemented in scikit-learn software,[53] was used to train optimal models. RF[50,51] randomly selects different features and training samples, generates many decision trees, and then average the results of these decision trees to perform the final classification. Following our test computations for different parameters of RF, the max depths for ORR/HER models were set as eight, and the numbers of trees were set as 500; while for OER model, the max depth was 7 with 1000 decision trees.

For the pretreatment of data, it is necessary to divide the training set and the cross-validation set. The training set was used to build the model, and the score of fitting for the test set would feedback to the model. The ratio of cross-validation sets

was tested and judged by the training scores and testing scores, and was finally set to around 0.85, 0.83, and 0.82 for ORR, OER, and HER, respectively. The 104 input data were randomly split into 88 training set and 16 validation set for ORR/HER models. For OER model, the small data set (with 26 data) for M-N-pyrrole systems was amplified three times (totally 156 input data), which was divided into 135 training data and 21 validation data. To lower the possibility of over-fitting, we reproduced the minor class with uniformly distributed random noises from -2% to 2% of the original data. Due to the large region of feature values, data normalization of all the features was performed before training. The training/testing score is the coefficient of determination ($R^2$) of the prediction, which is defined as $R^2 = 1 - \Sigma(y\_true - y\_pred)^2 / \Sigma(y\_true - y\_average)^2$. The mean squared error (MSE) represents the mean difference between the predicted values and the real values, defined as $MSE = 1/n\ \Sigma(y\_true - y\_pred)^2$.

DFT computations were carried out for optimization and frequency computations of the 19 top ML-recommended $M@N_xC_y$ SACs and adsorbates with *H, *O, *OH, and *OOH. These computations employed an all-electron method within a generalized gradient approximation (GGA) for the exchange-correlation term, as implemented in the DMol$^3$ code,[54,55] in which the double numerical plus polarization (DNP) basis set and Perdew, Burke and Ernzerhof (PBE) functional were adopted[56]. Self-consistent field (SCF) computations were performed with a convergence criterion of $10^{-6}$ a.u. To prevent artificial interactions between periodic images, we

applied a vacuum space of at least 15 Å in the perpendicular direction of the two-dimensional (2D) layer.

The changes of Gibbs free energy ($\Delta G$) for each elementary step along the ORR/OER were evaluated using the computational hydrogen electrode (CHE) model developed by Nørskov and co-workers.[57] The computed electronic energy was converted into Gibbs free energy by adding zero-point energy. The energy of the triplet $O_2$ molecule cannot be accurately computed by DFT method,[58,59] thus, its free energy was computed relative to the free energies of $H_2O(l)$ and $H_2(g)$. The chemical potential of the $H^+/e^-$ pair is equal to half of the gas-phase $H_2$ at standard hydrogen electrode (SHE) conditions.

## 3. RESULTS AND DISCUSSION

### 3.1. Graphene-supported SACs for ORR/OER/HER used to construct ML models

Three types of graphene-supported SACs (Figure 2) were chosen as representatives of the M-N-C electrocatalysts, in which the central metal atom is (1) at the single vacancy with three carbon atoms (M@$C_3$); (2) at the double vacancy with four nitrogen/carbon atoms (M@pyridine-$N_4$ /M@$C_4$); and (3) coordinated with four pyrrole nitrogen atoms (M@pyrrole-$N_4$). The 28 transition metals (except for Hg, La, and Ac) were employed as the central metal atoms. The data for training and validation are the limiting potentials and features of these SACs[21].

RF[50,51] is a widely used algorithm for both regression and classification problems, which randomly selects different features and training samples, generates many decision trees, and then average these decision trees to perform the final results. RF can not only depict the underlying patent of a complicated problem, but also provide feature importance for different features after training, which cannot be obtained by many other algorithms. Moreover, compared with Decision Tree, RF greatly improves the accuracy of models and avoids the easily attacked characteristic of DT; therefore, we employed the RF algorithm to train three models.

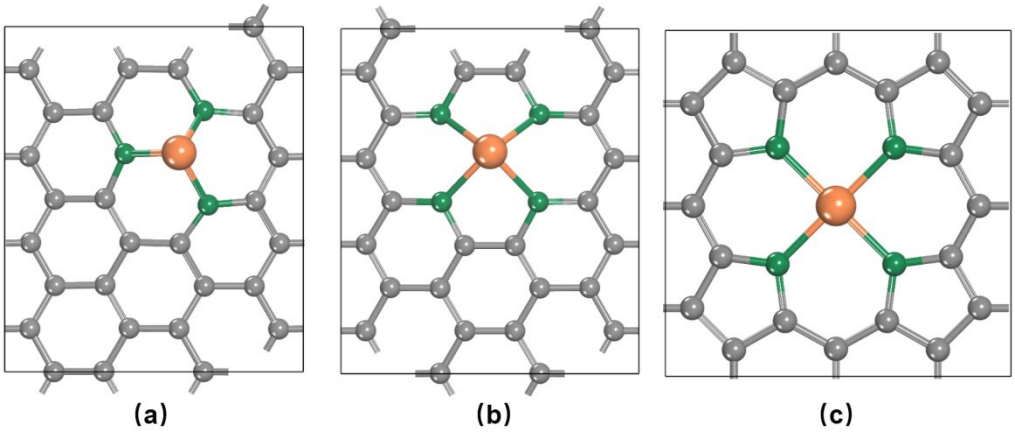



**Figure 2.** Schematic structures of graphene-supported SACs used to build up the ML models. (a) Single vacancy with three carbon atoms. (b) Double vacancy with four nitrogen/carbon atoms. (c) Four pyridine nitrogen atoms. The orange, green, and gray balls represent TM, neighboring N/C, and other C atoms, respectively.

### 3.2 Training ML Model for ORR

**Feature Set for ORR.** Obtaining an appropriate feature set with a compact composition and sufficient information is the most significant step for a ML process, since the quality choice of features determines the highest accuracy that a model can achieve. Focusing on intrinsic properties, we selected five features for ORR: the electron numbers of *d* orbital (*d*), the oxide formation enthalpy ($H_{of}$),[60,41] Pauling electronegativity of the center metal atom ($E_m$), the sum of Pauling electronegativity of surrounding atoms ($E_s$), and the average of the pKa values of the surrounding atoms (*pKa*).

The first feature, the electron numbers of *d* orbital (*d*), highly relates to the reactivity because the central atom would lose some ability to donate or accept electrons when its *d* orbital is filled with either too few or too many electrons.

The second feature, the oxide formation enthalpy ($H_{of}$) of a single atom, firmly connects with the ability of the metal atom to react with oxygen,[61] which can be obtained by using the following equation:

$$\Delta H_{of} = \Delta H_{sub} - H_{of,\ bulk},$$

where $\Delta H_{sub}$ is the experimentally determined cohesive energy of the bulk metal structure, and $\Delta H_{of,\ bulk}$ is the formation enthalpy of the metal's most stable oxide relative to the bulk metal and $O_2$. Moreover, the oxide formation enthalpy would not change along with different systems, which demonstrates that the oxide formation enthalpy values can be used directly in future studies.

The third and fourth features are the electronegativity values of the central metal atom and its surrounding atoms, respectively. The higher the associated electronegativity value, the more an atom or a substituent group can attract electrons towards itself. Notably, the electronegativity of an atom is influenced by its intrinsic property and the circumstance, thus we included both the electronegativity of the central atom ($E_m$) and the sum of electronegativity of surrounding atoms ($E_s$) into the feature set.

The last feature, the average of the p*Ka* values of the surrounding atoms (*pKa*), is introduced, to our best knowledge for the first time, as a feature for catalytic activity. Note that $E_s$ alone cannot well describe the environment of the neighboring atoms of the metal center, such as pyridine-4N and pyrrole-4N; while *pKa* can represent the activity of the neighboring N/C atoms in the M@$N_xC_y$ subunits. The p*Ka* values for pyridine-N, pyrrole-N, $C_6$-C, and $C_5$-C atoms are 5.25, 16.5, 44.0, and 15.0 (based on the pyridine, pyrrole, benzene, and cyclopentadiene molecules).

Among these five features, the first three (*d,* $H_{of}$, and $e_m$) are the intrinsic characteristics of the central metal atom itself, while the last two ($e_s$ and *pKa*) describe the environment of the central atom. Through these five descriptors, we can outline the connection of the physical properties and the catalysis performance of M-N-C for ORR.

**Performance of the ML Model for ORR.** Using the aforementioned five features, we successfully built up an ML model with excellent performance: the training and

test scores are 0.97 and 0.95, respectively. The ML predicted $U_L$ values well agree with our previous DFT data (Figure 3a): the average error (0.013 V) approaches the accuracy of DFT computations, and the mean square error (MSE, 0.027) is rather small, which verify the great performance of our ORR model. This ML model can also directly predict the limiting potentials of other related SACs, which will be discussed in Section 3.4. All the well-trained models and training/test data can be obtained by the link in Supporting Information.

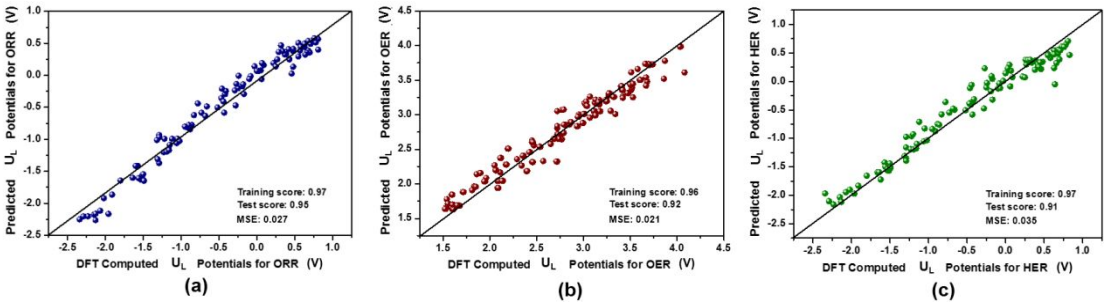



**Figure 3.** Comparison between DFT and ML predicted limiting potential ($U_L$) values of (a) ORR, (b) OER, and (c) HER, where both training and testing data points are presented.

To encode the significance of different properties towards $U_L$, we compared the importance of these five features (Figure 4). We found what dominate are the three features intrinsic to the central metal atom alone ($d$, $H_{of}$, and $E_m$), thus the type of central atoms is the most important factor for the catalytic performance. Among these three features, the number of d electrons ($d$) has the largest feature importance of

0.498, followed by the oxygen formation enthalpy ($H_{of}$) and the electronegativity ($E_m$) with the feature importance of 0.173 and 0.155, respectively. The sum of electronegativity of surrounding atoms ($E_s$) and the average of pKa ($pKa$) have less but still critical importance with the same feature importance of 0.087. The high training and test scores, as well as the rather high feature importance, verify that our ML model for ORR is reliable and our feature set is a good choice.

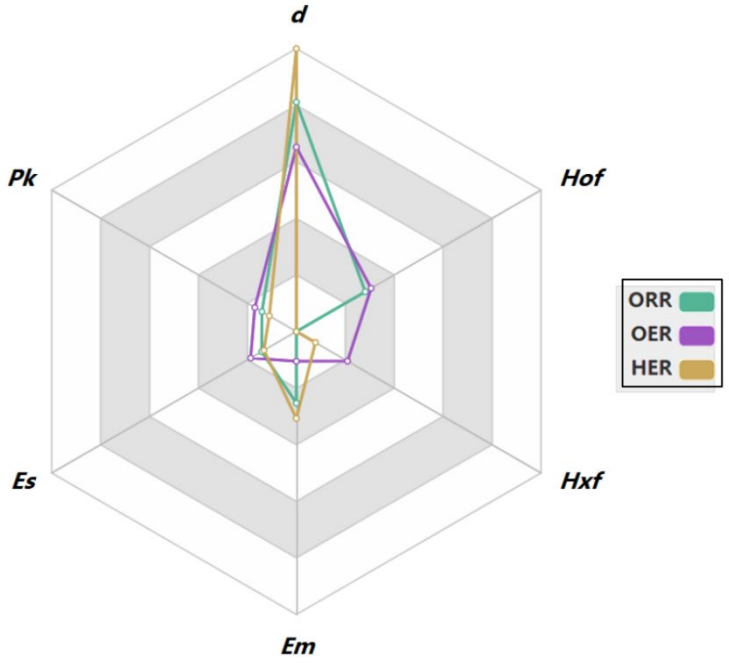



**Figure 4**. Radar chart on the feature importance of six features ($d$, $pKa$, $E_s$, $E_m$, $H_{xf}$, and $H_{of}$) in the ML models for ORR, OER, and HER.

### 3.3. Training ML Models for OER and HER

**Feature Set for OER.** The ML model for OER can only be trained well after involving another feature, namely the hydride formation enthalpy ($H_{xf}$) of a single

atom, besides the five features we used for ORR. Our newly introduced feature, $H_{xf}$, is conceptually similar to $H_{of}$, which reflects the hydrogen affinity of the central metal atom. The $H_{xf}$ value can be obtained by checking the energy difference between the most stable metal hydride and the most stable pure substances ($H_2$ gas and the metal bulk). The equation for calculating $H_{xf}$ is:

$$\Delta H_{xf} = \Delta H_{sub} - H_{xf,\ bulk},$$

where $\Delta H_{sub}$ is the experimentally determined cohesive energy of the bulk metal structure, $\Delta H_{xf,bulk}$ is the formation enthalpy of the metal's most stable hydride relative to the bulk metal and $H_2$.

Note that the majority of transition metals have stable compounds in nature, and in recent years many stable 2D metal hydrides (eg. Mn, Cu, Zn, Mo, Ru, Cd, W, Os, Ir, and Pt) have been achieved.[62,63] From the stable 2D metal hydride structures reported by Zhou *et al.*[62], we extracted those with the most negative formation energy for each transition metal. However, for two elements, namely Au and Ag, no stable compounds and 2D metal hydrides exist, thus, the M-N-C systems with Au and Ag were not covered in the ML models[62-64] (totally 26 transition metals were considered here).

**Performance of the ML model for OER.** Based on six features, the random forest model for OER also gained satisfactory scores for the training (0.96) and the test sets (0.92). The ML predicted limiting potentials have a good linear correlation with the

previous DFT results (Figure 3b). The rather low average error (0.005 V) and MSE (0.021) confirm the high performance of our ML model.

By examining the importance of six intrinsic features (Figure 4), we can disclose the relationships between descriptors and catalytic performance of the OER SACs under investigation. The number of d electrons ($d$) remains as the most significant feature for OER, but its importance (0.400) is slightly lower than that for ORR (0.498). The hydride ($H_{xf}$) and oxide formation enthalpies ($H_{of}$) are the second and the third vital features for OER (with the importance of 0.284 and 0.187, respectively). Compared with ORR ($H_{of}$ and $E_m$ ranked as the second and third features with the importance of 0.173 and 0.155, respectively), the formation enthalpy of hydrides ($H_{xf}$) has a much larger weight for estimating the limiting potentials of OER. The sum of electronegativity of surrounding atoms ($E_s$) and the average of Pka also play indispensable roles in the OER model with feature importance of 0.115 and 0.104, while the Pauling electronegativity of central metal has the lowest feature importance of 0.064.

**Feature set and performance of the ML model for HER.** To build up the ML models for HER, we used the five descriptors, including $H_{xf}$ but not $H_{of}$. This selection can be easily understood since only the strength of hydrogen bonding with central metal atom, rather than that of oxygen bonding, affects the catalytic performance for HER. Our ML model achieved scores of 0.97 and 0.91 for training and test sets,

respectively, and the average error and MSE are only 0.020 V and 0.035 for both sets (Figure 3c).

For feature importance exploration, we found that the same as ORR and OER, the number of d electrons (*d*) has the highest feature importance (of 0.614) for HER (Figure 4), which verifies again the unique role of the central transition metal for SACs. Ranked by the importance for HER, the electronegativity of central atoms ($E_m$) is in the second place (0.189), followed by the sum of Pauling electronegativity of surrounding atoms ($E_s$, 0.081), the average of the pKa values of the surrounding atoms (*pKa*, 0.068), and the hydride formation enthalpy ($H_{xf}$, 0.048).

To summarize, the intrinsic features we selected can be easily obtained and employed for the feature studies for SACs. The high scores of our three ML models demonstrate that the compact five/six features we extracted have sufficient information to describe and distinguish these SACs structures, which make us confident in using these ML models to predict the activity of other related SCAs towards these three reactions.

### 3.4. Predicting the limiting potentials of other graphene-supported SACs ($M@N_xC_y$) by ML models

What inspiring most is that ML techniques can quickly identify new materials with desired properties, thus significantly accelerating materials discovery and design. In this section, we will use the aforementioned well trained ML models to estimate

the catalytic performance (by predicting the limiting potentials) of many more graphene-supported SACs with similar structural configurations.

It is known that the coordination environments of the central metal atom are important factors for the performance of SACs. However, no systematic investigation on the catalytic performance of enormous graphene-supported SACs towards ORR/OER/HER has been performed. Thus, we constructed 260 $M@N_xC_y$ SACs containing 26 transition metals and with 10 types of metal coordination environments, including $M@N_1C_2$, $M@N_2C_1$, $M@N_3$, M@pyridine-$N_1C_3$, M@pyridine-$N_2C_2$, M@pyridine-$N_3C_1$, M@pyrrole-$C_4$ M@pyrrole-$N_1C_3$, M@pyrrole-$N_2C_2$, and M@pyridine-$N_3C_1$.

In total, 1560 descriptors (six each) were used as input of the ML models to predict the limiting potentials of these 260 $M@N_xC_y$ SACs towards ORR/OER/HER. The whole prediction process took only several seconds, and their limiting potentials were obtained without any geometry optimization and energy computations. Among these 260 SACs, six catalysts show high $U_L$ values (~0.55 V) for ORR (corresponding to overpotential $\eta$ ~0.68 V), four achieve low $U_L$ values for OER (~1.64 V, $\eta$= ~0.41 V), and 12 have close-to-zero $U_L$ values for HER ($\eta$ < 0.04 V) (for details, see Table S1 in Supporting Information).

We then examined the common characteristics of the top 3/4/12 M-N-Cs catalysts (ranked by overpotentials). The ML-recommended M-N-Cs for ORR are $Fe@N_3C_1$, Fe@pyrrole-$N_2C_2$, and Fe@pyrrole-$N_1C_3$; for OER are Ir@pyridine-$N_3C_1$,

Ir@pyrrole-$N_3C_1$, Ir@pyrrole-$N_2C_2$, and Ir@pyrrole-$N_3C_1$; for HER are Mo@pyridine-$N_2C_2$, Zn@$N_1C_2$, Pd@$N_2C_1$, Re@pyridine-$N_2C_2$, Rh@$N_3$, Pt@$N_2C_1$, Rh@$N_1C_2$, Zn@$N_3$, Ni@pyridine-$N_3C_1$, Re@pyrrole-$N_4$, Re@pyrrole-$N_1C_3$, Ni@pyrrole-$N_2C_2$, and Ni@pyrrole-$N_3C_1$. Interestingly, for ORR all the recommended SACs contain Fe as the metal center, while for OER, Ir serves as the best active site. These data clearly demonstrate that Fe and Ir are eminent active centers for ORR and OER, respectively, which echo with previous experimental and theoretical findings[65-69].

To examine the accuracy of ML models, we computed the limiting potentials of these top M-N-Cs for ORR/OER/HER (three for ORR, four for OER, and 12 for HER) by DFT computations. The SACs with the highest limiting potentials (lowest overpotentials) for ORR are Fe@pyrrole-$N_1C_3$ and Fe@pyrrole-$N_2C_2$, i.e. Fe@pyrrole-$N_1C_3$ has the highest $U_L$ value of 0.61 V (Figure 5a, $\eta$ = 0.62 V), and Fe@pyrrole-$N_2C_2$ has the second highest $U_L$ value of 0.55 V (Figure 5a, $\eta$ = 0.68 V). For SACs towards OER, Ir@pyridine-$N_3C_1$ has the lowest limiting potential ($U_L$ = 1.51 V, $\eta$ = 0.28 V) (Figure 5b), followed by Ir@pyrrole-$N_2C_2$ ($U_L$ =1.55 V, $\eta$ = 0.22 V). These two ML-selected OER catalysts perform better than the generally regarded best OER catalysts, such as $RuO_2$ ($U_L$ = 1.60 V)[70], $IrO_2$ ($U_L$ = 1.88 V)[21], and doped carbon materials (P-doped graphdiyne ($U_L$ = 1.58 V)[70], and previous studied M@N/C system ($U_L$ > 1.52 V)[21]. The average errors of the ML predicted limiting potential values (relative to the DFT values) of the three ORR and four OER catalysts are 0.09

and 0.06 V, and the corresponding MSE values are 0.11 and 0.03, respectively, which demonstrate the robustness and accuracy of our ML models.

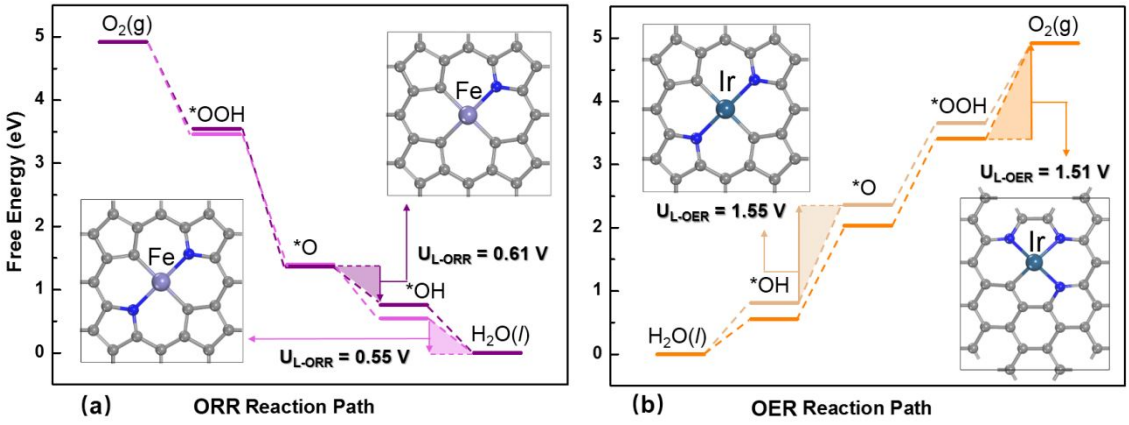



**Figure 5.** Free-energy diagrams and geometries of prominent M@$N_xC_y$ SACs for (a) ORR and (b) OER at zero electrode potential. The rate-determining steps are highlighted by shades; and the blue and gray balls stand for nitrogen and carbon atoms, respectively.

When compared with the DFT computed values, the average error of the 12 ML-predicted limiting potentials is 0.07 V, and the corresponding MSE value is 0.18, which demonstrate the rather high reliability of our ML models. The best catalyst for HER is Ni@pyridine-$N_3C_1$ with a $U_L$ value of only 0.003 V, which is very close to 0 V (Figure 6, $\eta$ = 0.003 V). Ni@pyridine-$N_3C_1$ also outperforms the conventional HER catalyst-Pt ($U_L$ of Pt is around -0.09 ~ 0.03)[71], previous studied Tc@$C_3$ ($U_L$ ~ -0.03 [72]), and has similar performance as Fe@$C_3$ ($U_L$ ~ 0.00) [72].

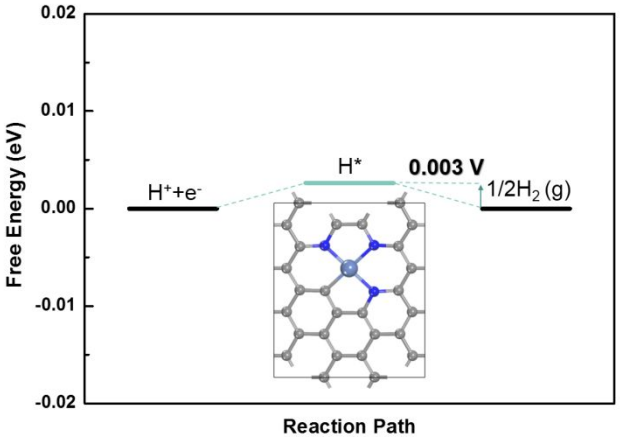



**Figure 6.** Free-energy diagram and geometry of the best-performed SAC for HER, Ni@pyridine-$N_3C_1$, at zero electrode potential. Blue ashes, navy, and gray balls stand for transition metal, nitrogen, and carbon atoms, respectively.

## 4. CONCLUSION

In summary, by machine learning techniques, we built up three models to describe the underlying pattern of physical properties and limiting potentials towards ORR/OER/HER of 104 graphene-supported SACs, and quantitatively unveiled the significance of various descriptors. Furthermore, we employed these ML models to directly predict the limiting potentials of 260 M@$N_xC_y$ SACs, and screened out the most promising ORR/OER/HER catalysts. The excellent catalytic performances of these ML-recommended SACs were verified by DFT computations, and the best candidates possess limiting potentials of 0.61, 1.51, and 0.003 V for ORR, OER, and HER. Especially, two OER catalysts (Ir@pyridine-$N_3C_1$, $U_L$=1.51 V; Ir@pyridine-$N_2C_2$, $U_L$=1.55 V) outperform the most commonly used noble metal

oxides ($RuO_2$ and $IrO_2$); and one HER catalyst (Ni@pyridine-$N_3C_1$, $U_L$=0.003 V) also outperform the commonly used noble metal (Pt).

By utilizing only a few easily available intrinsic physical features of M-N-Cs, our ML models can well predict the limiting potentials of graphene-supported SACs towards ORR/OER/HER. Without any geometry optimization, total energy calculation, or examining reaction pathways, this ML process takes only seconds, but can dramatically narrow down the candidate list of M-N-C SACs. This strategy can be used to screen and design other electrochemical catalysts, such as towards nitrogen reduction reactions and $CO_2$ reduction reactions. Directly predicting catalytic performance of electrocatalysts from the easily obtainable parameters of catalysts is bringing us a revolutionary approach for future catalysts design, and will dramatically accelerate the discovery of more efficient catalysts towards important chemical processes in the very near future.

## ASSOCIATED CONTENT

### Supporting Information

Data pretreatment; Basic information for DFT computations for catalysis performance; Github website link for the training data, prediction data and the well-trained models; The geometries of ML-selected best SACs for ORR/OER/HER; The limiting potentials of 260 other graphene-supported SACs (M@$N_xC_y$) towards ORR/OER/HER predicted by ML models.

## AUTHOR INFORMATION

**Corresponding Author**

* zhongfangchen@gmail.com

**Notes**

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work is supported by the National Science Foundation-Centers of Research Excellence in Science and Technology (NSF-CREST Center) for Innovation, Research and Education in Environmental Nanotechnology (CIRE2N) (Grant No. HRD-1736093).

## REFERENCES

(1) Amiinu, I. S.; Liu, X.; Pu, Z.; Li, W.; Li, Q.; Zhang, J.; Tang, H.; Zhang, H.; Mu, S., *Adv. Funct. Mater.* **2018,** *28* (5), 1704638.

(2) Li, Y.; Lu, J., *ACS Energy Lett.* **2017,** *2* (6), 1370-1377.

(3) Luo, J.; Im, J.-H.; Mayer, M. T.; Schreier, M.; Nazeeruddin, M. K.; Park, N.-G.; Tilley, S. D.; Fan, H. J.; Grätzel, M., *Science* **2014,** *345* (6204), 1593-1596.

(4) Liu, L.; Zeng, G.; Chen, J.; Bi, L.; Dai, L.; Wen, Z., *Nano Energy* **2018,** *49*, 393-402.

(5) Chung, H. T.; Cullen, D. A.; Higgins, D.; Sneed, B. T.; Holby, E. F.; More, K. L.; Zelenay, P., *Science* **2017,** *357* (6350), 479-484.

(6) Santoro, C.; Serov, A.; Gokhale, R.; Rojas-Carbonell, S.; Stariha, L.; Gordon, J.; Artyushkova, K.; Atanassov, P., *Appl. Catal. B-Environ.* **2017,** *205*, 24-33.

(7) Chen, C.; Kang, Y.; Huo, Z.; Zhu, Z.; Huang, W.; Xin, H. L.; Snyder, J. D.; Li, D.; Herron, J. A.; Mavrikakis, M., *Science* **2014,** *343* (6177), 1339-1343.

(8) Wang, Y.-J.; Zhao, N.; Fang, B.; Li, H.; Bi, X. T.; Wang, H., *Chem. Rev.* **2015,** *115* (9), 3433-3467.

(9) Zhang, B.; Fan, T.; Xie, N.; Nie, G.; Zhang, H., *Adv. Sci.* **2019**, 1901787.

(10) He, T.; Matta, S. K.; Will, G.; Du, A., *Small Methods* **2019**, 1800419.

(11) Liu, X.; Jiao, Y.; Zheng, Y.; Jaroniec, M.; Qiao, S.-Z., *J. Am. Chem. Soc.* 2019. *141*(24), 9664-9672.

(12) Alexopoulos, K.; Wang, Y.; Vlachos, D. G., *ACS Catal.* **2019,** *9*, 5002-5010.

(13) Bakandritsos, A.; Kadam, R. G.; Kumar, P.; Zoppellaro, G.; Medved', M.; Tuček, J.; Montini, T.; Tomanec, O.; Andrýsková, P.; Drahoš, B., *Adv. Mater.* **2019,** *31* (17), 1970125.

(14) Yang, X.-F.; Wang, A.; Qiao, B.; Li, J.; Liu, J.; Zhang, T., *Acc. Chem. Res.* **2013,** *46* (8), 1740-1748.

(15) Qiao, B.; Wang, A.; Yang, X.; Allard, L. F.; Jiang, Z.; Cui, Y.; Liu, J.; Li, J.; Zhang, T., *Nat. Chem.* **2011,** *3* (8), 634.

(16) Lefèvre, M.; Proietti, E.; Jaouen, F.; Dodelet, J.-P., *Science* **2009,** *324* (5923), 71-74.

(17) Chen, Y.; Ji, S.; Wang, Y.; Dong, J.; Chen, W.; Li, Z.; Shen, R.; Zheng, L.; Zhuang, Z.; Wang, D., *Angew. Chem. Inter. Ed.* **2017,** *56* (24), 6937-6941.

(18) Chen, W.; Pei, J.; He, C. T.; Wan, J.; Ren, H.; Zhu, Y.; Wang, Y.; Dong, J.; Tian, S.; Cheong, W. C., *Angew. Chem. Inter. Ed.* **2017,** *56* (50), 16086-16090.

(19) Costentin, C.; Savéant, J.-M., *Nat. Rev. Chem.* **2017,** *1* (11), 0087.

(20) Zagal, J. H.; Griveau, S.; Silva, J. F.; Nyokong, T.; Bedioui, F., *Coord. Chem. Rev.* **2010,** *254* (23-24), 2755-2791.

(21) Xu, H.; Cheng, D.; Cao, D.; Zeng, X. C., *Nat. Catal.* **2018,** *1* (5), 339.

(22) He, X.; He, Q.; Deng, Y.; Peng, M.; Chen, H.; Zhang, Y.; Yao, S.; Zhang, M.; Xiao, D.; Ma, D., *Nat. Commun.* **2019,** *10* (1), 1-9.

(23) Guo, X.; Gu, J.; Hu, X.; Zhang, S.; Chen, Z.; Huang, S., *Catal. Today* **2019**, 10.1016/j.cattod.2019.06.014.

(24) Lin, C. Y.; Zhang, L.; Zhao, Z.; Xia, Z., *Adv. Mater.* **2017,** *29* (17), 1606635.

(25) Greeley, J.; Nørskov, J. K., *Electrochim. Acta* **2007,** *52* (19), 5829-5836.

(26) Trevor, H.; Robert, T.; JH, F., The elements of statistical learning: data mining, inference, and prediction. New York, NY: Springer: **2009**, 27(2), 83-85.

(27) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C., *npj Comput. Mater.* **2016,** *2*, 16028.

(28) Sahu, H.; Rao, W.; Troisi, A.; Ma, H., *Adv. Ener. Mater.* **2018,** *8* (24), 1801032.

(29) Lu, S.; Zhou, Q.; Ouyang, Y.; Guo, Y.; Li, Q.; Wang, J., *Nat. Commun.* **2018,** *9* (1), 3405.

(30) Li, W.; Field, K. G.; Morgan, D., *npj Comput. Mater.* **2018,** *4* (1), 36.

(31) Tshitoyan, V.; Dagdelen, J.; Weston, L.; Dunn, A.; Rong, Z.; Kononova, O.; Persson, K. A.; Ceder, G.; Jain, A., *Nature* **2019,** *571* (7763), 95.

(32) Severson, K. A.; Attia, P. M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M. H.; Aykol, M.; Herring, P. K.; Fraggedakis, D., *Nat. Ener.* **2019,** *4* (5), 383.

(33) Zhou, Q.; Tang, P.; Liu, S.; Pan, J.; Yan, Q.; Zhang, S.-C., *Proc. Natl. Acad. Sci.* **2018,** *115* (28), E6411-E6417.

(34) Mansouri Tehrani, A.; Oliynyk, A. O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T. D.; Brgoch, J., *J. Am. Chem. Soc.* **2018,** *140* (31), 9844-9853.

(35) Raccuglia, P.; Elbert, K. C.; Adler, P. D.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J., *Nature* **2016,** *533* (7601), 73.

(36) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C., *npj Comput. Mater.* **2017,** *3* (1), 54.

(37) Correa-Baena, J.-P.; Hippalgaonkar, K.; van Duren, J.; Jaffer, S.; Chandrasekhar, V. R.; Stevanovic, V.; Wadia, C.; Guha, S.; Buonassisi, T., *Joule* **2018,** *2*(8), 1410.

(38) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A., *Nature* **2018,** *559* (7715), 547.

(39) Zhang, Y.; Ling, C., *npj Comput. Mater.* **2018,** *4* (1), 25.

(40) Askerka, M.; Li, Z.; Lempen, M.; Liu, Y.; Johnston, A.; Saidaminov, M. I.; Zajacz, Z.; Sargent, E. H., *J. Am. Chem. Soc.* **2019,** *141* (8), 3682-3690.

(41) O'Connor, N. J.; Jonayat, A.; Janik, M. J.; Senftle, T. P., *Nat. Catal.* **2018,** *1* (7), 531.

(42) Li, H.; Zhang, Z.; Liu, Z., *Catalysts* **2017,** *7* (10), 306.

(43) Tian, Z.; Priest, C.; Chen, L., *Adv. Theory Simul.* **2018,** *1* (5), 1800004.

(44) Gasper, R.; Shi, H.; Ramasubramaniam, A., *J. Phys. Chem. C* **2017,** *121* (10), 5612-5619.

(45) Jinnouchi, R.; Hirata, H.; Asahi, R., *J. Phys. Chem. C* **2017,** *121* (47), 26397-26405.

(46) Chen, Y.; Huang, Y.; Cheng, T.; Goddard, W. A., *J. Am. Chem. Soc.* **2019**. *141*(29), 11651-11657

(47) Bai, Y.; Wilbraham, L.; Slater, B. J.; Zwijnenburg, M. A.; Sprick, R. S.; Cooper, A. I., *J. Am. Chem. Soc.* **2019**. *141*(22), 9063-9071

(48) Tibshirani, R., *J. R. Stat. Soc. B* **1996,** *58* (1), 267-288.

(49) Ma, X.; Li, Z.; Achenie, L. E.; Xin, H., *J. Phys. Chem. Lett.* **2015,** *6* (18), 3528-3533.

(50) Liaw, A.; Wiener, M., *R. News* **2002,** *2* (3), 18-22.

(51) Díaz-Uriarte, R.; De Andres, S. A., *BMC Bioinformatics* **2006,** *7* (1), 3.

(52) Rokach, L.; Maimon, O., Classification trees. In *Data mining and knowledge discovery handbook*, Springer: 2009; pp 149-174.

(53) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V., *J. Mach. Learn. Res.* **2011,** *12* (Oct), 2825-2830.

(54) Delley, B., *J Chem Phys* **1990,** *92* (1), 508-517.

(55) Delley, B., *J. Chem. Phys.* **2000,** *113* (18), 7756-7764.

(56) Perdew, J. P.; Burke, K.; Ernzerhof, M., *Phys. Rev. Lett.* **1996,** *77* (18), 3865.

(57) Nørskov, J. K.; Rossmeisl, J.; Logadottir, A.; Lindqvist, L.; Kitchin, J. R.; Bligaard, T.; Jonsson, H., *J. Phys. Chem. B* **2004,** *108* (46), 17886-17892.

(58) Jones, R. O.; Gunnarsson, O., *Rev. Mod. Phys.* **1989,** *61* (3), 689.

(59) Kurth, S.; Perdew, J. P.; Blaha, P., *Int. J. Quantum Chem.* **1999,** *75* (4 - 5), 889-909.

(60) Campbell, C. T.; Sellers, J. R., *Faraday Discuss.* **2013,** *162*, 9-30.

(61) Guo, J.; Li, Y.; Cheng, Y.; Dai, L.; Xiang, Z., *ACS Nano* **2017,** *11* (8), 8379-8386.

(62) Zhou, X.; Hang, Y.; Liu, L.; Zhang, Z.; Guo, W., *J. Am. Chem. Soc.* **2019**, *141*(19), 7899-7905

(63) Wu, Q.; Zhang, Y.; Zhou, Q.; Wang, J.; Zeng, X. C., *J. Phys. Chem. Lett.* **2018,** *9* (15), 4260-4266.

(64) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G., *Apl. Mater.* **2013,** *1* (1), 011002.

(65) Lin, L.; Zhu, Q.; Xu, A.-W., *J. Am. Chem. Soc.* **2014,** *136* (31), 11027-11033.

(66) Liang, J.; Zhou, R. F.; Chen, X. M.; Tang, Y. H.; Qiao, S. Z., *Adv. Mater.* **2014,** *26* (35), 6074-6079.

(67) Ao, X.; Zhang, W.; Li, Z.; Li, J.-G.; Soule, L.; Huang, X.; Chiang, W.-H.; Chen, H. M.; Wang, C.; Liu, M., *ACS Nano* **2019**, 10.1021/acsnano.9b05913

(68) Seh, Z. W.; Kibsgaard, J.; Dickens, C. F.; Chorkendorff, I.; Nørskov, J. K.; Jaramillo, T. F., *Science* **2017,** *355* (6321), eaad4998.

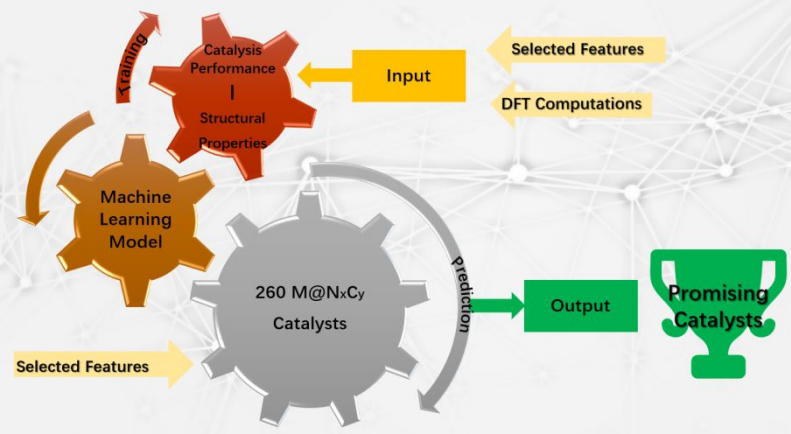
(69) Guo, J.; Huo, J.; Liu, Y.; Wu, W.; Wang, Y.; Wu, M.; Liu, H.; Wang, G., *Small Methods* **2019**, 1900159.

(70) Gu, J.; Magagula, S.; Zhao, J.; Chen, Z., *Small Methods* **2019**, 1800550.

(71) Nørskov, J. K.; Bligaard, T.; Logadottir, A.; Kitchin, J.; Chen, J. G.; Pandelov, S.; Stimming, U., *J. Electrochem. Soc.* **2005,** *152* (3), J23-J26.

(72)Choi, W. I.; Wood, B. C.; Schwegler, E.; Ogitsu, T., *Adv. Ener. Mater.* **2015,** *5* (23), 1501423.

**TOC**


Training
Catalysis Performance
Structural Properties
Input
Selected Features
DFT Computations
Machine Learning Model
260 M@NxCy Catalysts
prediction
Output
Promising Catalysts
Selected Features