

# **Screening of Highly Active Hydrogen Evolution Reaction Catalysts: Comparative Analysis of Typical Machine Learning Methods**

Jordan Mershimer.

Chemistry Department, Warren Wilson College.

Yuemei Zhang, Ph.D.

Kim Borges, Ph.D.

Langdon Martin, Ph.D.

Fall 2022.

## **Abstract**

Hydrogen evolution reaction (HER) catalysts play a critical role in renewable energy technologies such as hydrogen production. Identifying novel catalysts with high efficiency and low cost is a significant challenge. In this study, we investigate the effectiveness of supervised machine learning techniques for screening HER catalysts.

We applied various machine learning models to predict the probability of a list of 300 transition metal borides, carbides, and nitrides being a highly efficient HER catalyst. Elemental descriptors that can be easily obtained by the periodic table for Metal (M) and carbon, nitrogen, boron (X) were primarily used. Materials with published Gibbs free energy of hydrogen adsorption ( $\Delta G_{H^*}$ ) values were utilized for our training data. We compared the different machine learning models based on root mean square error (RMSE), speed of model generation, and speed of model prediction.

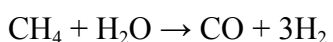
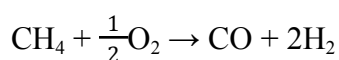
Our results showed that the Random Forest Regression model produced the lowest testing RMSE and was chosen to examine the 300 materials, Out of which 29 materials were predicted to be high-performance HER catalysts. Our approach can efficiently identify promising HER catalysts with high accuracy, providing guidance for further experimental and theoretical investigations.

## **Introduction**

If renewable energy is to eclipse fossil-fuels in energy production there are some obstacles that will need to be overcome. The three most prominent problems are those of intermittency, variability, and maneuverability.<sup>1</sup> Renewable energy is flow-limited, meaning that energy is only generated when the sun is shining, the wind is blowing, or the river is flowing. In comparison, fossil-fuels are not flow-limited: you can burn coal or natural gas so long as you

have reserves of it. Energy demand across time is not consistent; there are peaks and troughs in demand across a single day or season. At night, the sun is not shining, which means solar panels are not generating electricity for the lights in your home. Renewable energy production is also geographically limited: solar power is most effective in sunny locations, wind turbines in windy locations, and hydroelectric in areas with rivers. To solve these problems we need a way to easily store and transport large amounts of energy. Lithium-ion batteries may seem like a sensible solution but they degrade overtime<sup>2</sup>, require rare earth minerals to manufacture, and could be negatively impacting the environment<sup>3-5</sup>. Hydrogen gas, a fuel source that produces only water when combusted, has the potential to change the climate of energy production<sup>6,7</sup> by allowing easy transport and storage of energy. We could also leverage existing gas pipelines across the world to transport hydrogen gas<sup>8</sup>. Hydrogen gas is already stored industrially in underground salt caverns<sup>9-11</sup>, and there is also the possibility of storing it as a highly compressed gas or liquid in cryogenic tanks<sup>12</sup>, or as a hydride salt<sup>12</sup>. However, the scale of production needed to implement hydrogen gas as a meaningful storage and transport solution is not economically feasible with current methods<sup>6</sup>. The most widely used method for hydrogen production is steam methane reforming (SMR)<sup>13</sup>, which uses an oxygen transport membrane (OTM) and a hydrogen transport membrane (HTM) to separate hydrogen gas from methane. The reactions are described by Reaction 1-3 and depicted visually by Figure 1.

Reaction 1-2 in the OTM Reactor,



Reaction 3 in the HTM Reactor,

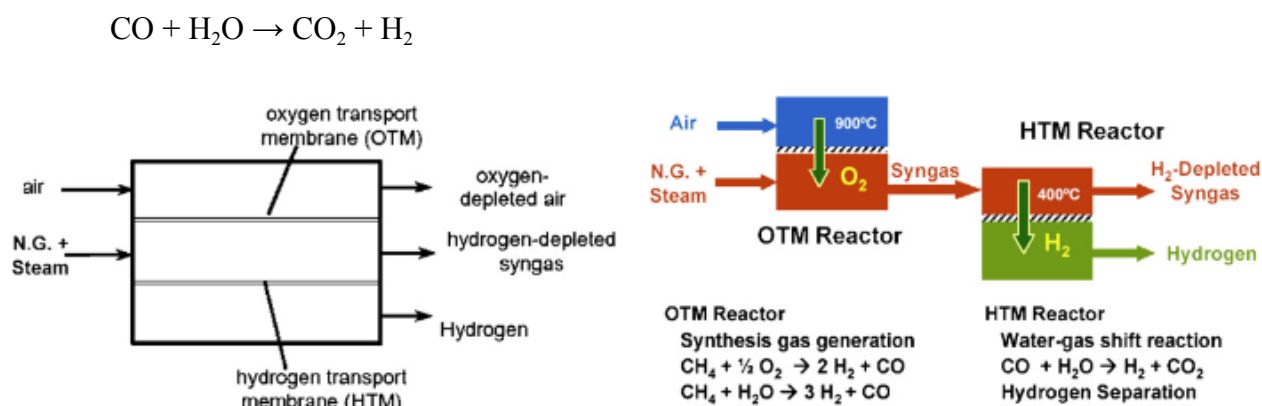
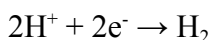


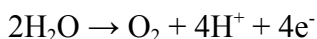
Figure 1. Integrated-membrane reactor separator.<sup>13</sup> Reprinted from ref 13. 2008 Barelli, L.

Currently, 80–85% of the world's total hydrogen gas is produced by SMR<sup>14</sup>. This process is very cheap, but it relies on fossil fuels and produces CO<sub>2</sub>. Water can be separated into oxygen and hydrogen gas through a process called electrolysis<sup>15</sup>. This process is depicted visually in Figure 2.<sup>16</sup> Water electrolysis is a promising alternative to SMR. During water electrolysis there are two half reactions: Hydrogen Evolution Reaction<sup>17</sup> (HER) and Oxygen Evolution Reaction (OER). The two half reaction mechanisms differ based on the pH of the media. The reactions are described in Reaction 4-7<sup>20</sup> and summarized visually in Figure 2.

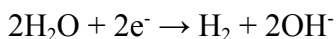
Reaction 4. Hydrogen Evolution Reaction in acidic media.



Reaction 5. Oxygen Evolution Reaction in acidic media.



Reaction 6. Hydrogen Evolution Reaction in alkaline media.



Reaction 7. Oxygen Evolution Reaction in alkaline media.



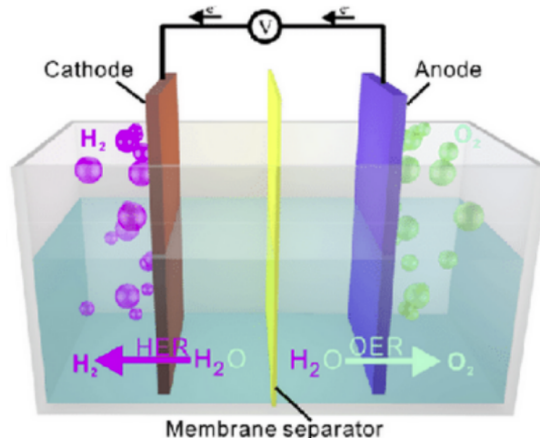


Figure 2. Hydrogen evolution reaction and oxygen evolution reaction of electrocatalytic water splitting.<sup>16</sup> Reprinted from ref 21. 2020 Zhu, J.

The minimum energy required for this reaction to occur is  $237 \text{ kJ mol}^{-1}$ . This process is not 100% efficient and requires an overpotential of energy for it to occur. We can cover the electrode in a catalyst that makes the conversion from electricity to H<sub>2</sub> gas more efficient and reduce the overpotential of the reaction.

There are many known HER catalysts and many are very effective such as noble metal-based catalysts like Pt, Ru, Rh, etc; however, these catalysts come with high material costs and are not very abundant on earth which make them unfavorable for use at scale. These materials serve as important benchmarks to compare against other catalysts.

When designing new HER catalysts we should aim for similar performance to Pt(111). We also know that highly acidic media allows for easy H<sup>+</sup> transfer; this would require a material that is stable under these conditions. It is also evident that the number of active sites available on the catalyst surface and the surface area will be important considerations for catalyst design.

In recent times there has been an increased interest in 2D materials as catalysts such as MoS<sub>2</sub><sup>22</sup>, MXenes<sup>23</sup>, and MBenes<sup>24</sup> which exhibit small HER overpotential and acceleration of

reaction rate. MXenes are materials composed of early transitional metal (M) carbides and nitrides (X), and MBenes are materials composed of transitional metal (M) from 3B, 4B, and 5B on the periodic table and boron (B). These 2D materials are promising alternatives to noble-metal-based catalysts. One advantage is that they have the ability to be deposited more uniformly on a cathode and have large surface area. Some disadvantages are that many have poor charge-transfer performance<sup>22</sup>, are unstable under ambient conditions, or have very few active sites. There is still much searching left to do regarding highly active HER catalysts.

There are hundreds of thousands of materials which could have potential as HER catalysts that are already synthesized and their structures have been deposited in material databases. But their properties and applications have not yet been fully explored. Many of them could be highly efficacious HER catalysts; however, the typical method of using density functional theory (DFT) calculations to calculate whether a material is an efficient catalyst is very time consuming. DFT is a computational method for modeling electronic structures of many-body systems such as solids and molecules based on electron density distribution. Calculations for a single material may take anywhere from a few hours to a few days, making it difficult to screen high numbers of materials. Machine learning may be a way around this problem.

Machine learning is the usage of computer algorithms that can self-improve by the use of data or training. Machine learning can screen hundreds or thousands of materials in a substantially smaller amount of time. Machine learning has already been applied successfully as a high-throughput screening tool for HER catalysts.<sup>24,25</sup> There have been models developed that can screen MBenes materials for HER catalyst activity,<sup>24</sup> and other models that can screen

bimetallic catalysts<sup>25</sup>. These applications typically target one specific group of 2D materials or use a single machine learning model.

There are many different approaches to machine learning; these are often referred to as models. These models intake tabular data known as data descriptors and use algorithms to output a result. The result can depend on the model used. The models belong to two types of machine learning methods, unsupervised learning and supervised learning.

Unsupervised learning uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data associations. The ability to discover similarities and differences in information make these algorithms the ideal solution for exploratory data analysis.<sup>26</sup>

Supervised learning is defined by the use of labeled datasets to train algorithms to classify data or predict outcomes accurately. Supervised learning uses two data-sets: training, and testing data-sets. As training data is fed into the model, it adjusts its weights until the model has been fitted appropriately. Finally, the model is used to predict outcomes of an entirely new data set known as the testing set, the true values of the testing set are known and are compared with the predicted values.<sup>27</sup> Supervised learning allows us to generate models that are capable of predicting outputs for future materials.

I plan to analyze and compare different supervised machine learning models as well as discern which data descriptors are most correlated with HER catalysts across MXenes<sup>23</sup> and MBenes<sup>24</sup>. The understanding of correlations between data descriptors and the differences between supervised learning models can then be used by others to inform further research on this topic.

## **Methods**

Hydrogen evolution reaction (HER) catalysts are used in the production of hydrogen gas to lower the energy required to split  $\text{H}_2\text{O}$  molecules. There are an exceedingly large number of potential HER catalysts and the time required by traditional means such as density functional theory (DFT) to calculate the properties of every potential candidate are enormous. Machine learning techniques may be used as effective screening tools to shrink the number of potential candidates to a more reasonable amount which can then be investigated by traditional means in the span of days instead of years.

We will be investigating multiple machine learning approaches to discern which is most efficacious for hydrogen evolution reaction (HER) catalyst screening.

Supervised learning techniques including Ordinary Least Squares, Gaussian process, Kernel Ridge, Ridge Cross Validation, Support Vectors, Decision Tree, Random Forest, and Multilayer perceptron were applied in our study..

We will test our supervised learning techniques and attempt to predict the probability of a generated list of 300 MXene and MBene materials being HER catalysts. Our 300 materials are composed of metals “Ti”, “Hf”, “V”, “Nb”, “Ta”, “Cr”, “Mo”, “W”, “Mn”, “Tc”, “Sc”, “Zr”, “Ru”, “Fe”, “Ni”, “Rh”, “Os”, “Co”, “Ir”, “Re”. As well as non-metals “B”, “N”, “C”. We will then compare our models based on root mean square error, speed of model generation and speed of model prediction.

We will be using entirely python for our codebase, the code is available for use under MIT license at [www.github.com/Jmersh/NSURS-HER-ML](https://www.github.com/Jmersh/NSURS-HER-ML). We will compile data descriptors using pymatgen’s periodic table module ([www.pymatgen.org](https://www.pymatgen.org)).



We often compare key characteristics such as their Gibbs energy ( $\Delta G$ ) and enthalpy ( $\Delta H$ ), for hydrogen electroadsorption to theoretical catalysts. From both theoretical and experimental perspectives the hydrogen adsorption free energy ( $\Delta G_{H^*}$ ) is the most important factor for describing the HER/OER activity of an electrode in acidic solutions.<sup>21</sup> Taking the most well-studied surfaces Pt(111) as an example, under pH 13, the Pt(111) surface has an optimal  $\Delta G_{H^*}$  close to zero. The closer to zero the  $\Delta G_{H^*}$  is the better performance of a catalyst. We will be using published Gibbs Free Energy adsorption values of the 300 materials for our labeled training data and using our supervised learning models to predict Gibbs Free Energy values for a set of materials as described earlier.

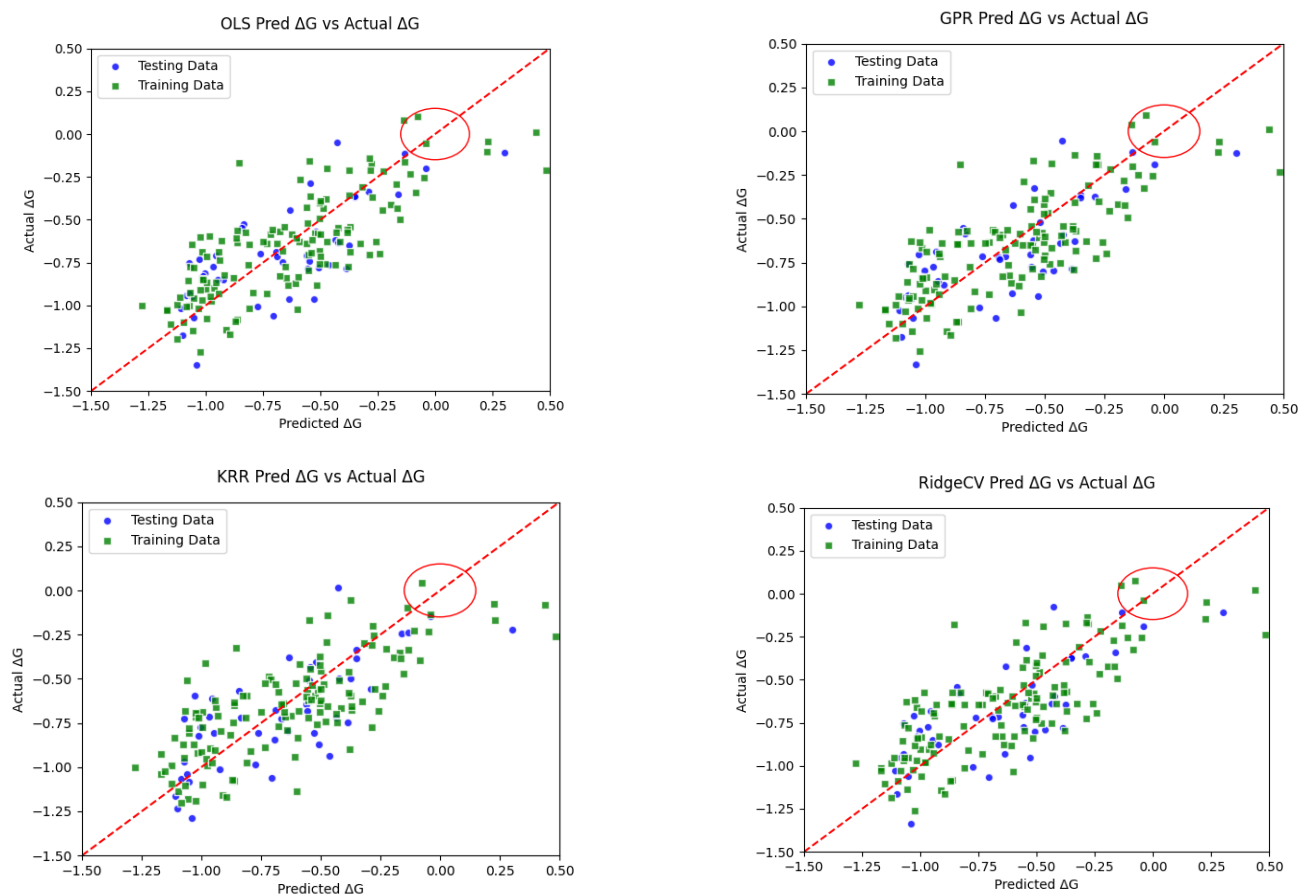
Table 1. List of descriptors used.

Elemental Descriptors	Material Descriptors
M/X Number	Gibbs Free energy
M/X Group	M/X Ratio
M/X Row	
M/X Mass	
M/X Density	
M/X Inter-atomic Distance	
M/X Covalent Radius	
M/X First Ionization Energy	
M/X Electron Affinity	

We are primarily focusing on elemental descriptors that can be easily obtained by the periodic table for the Metal (M) and carbon, nitrogen, boron (X). If we can generate accurate models using only these descriptors it will be extremely efficient and require no additional DFT or other calculations.

## Results

Figure 3 are plots of the predicted vs actual Gibbs free energy changes between the testing and training data using different models. The blue circles represent the testing data and the green squares represent the training data. The dashed red line serves as a reference between the predicted and actual Gibbs free energy change, the closer the squares or circles are to the line represents the model's increased accuracy, this can also allow us to detect abnormalities in the model such as overfitting. The red circle represents our zone of interest for Gibbs free energy change values, which is  $\pm 0.15$  eV.



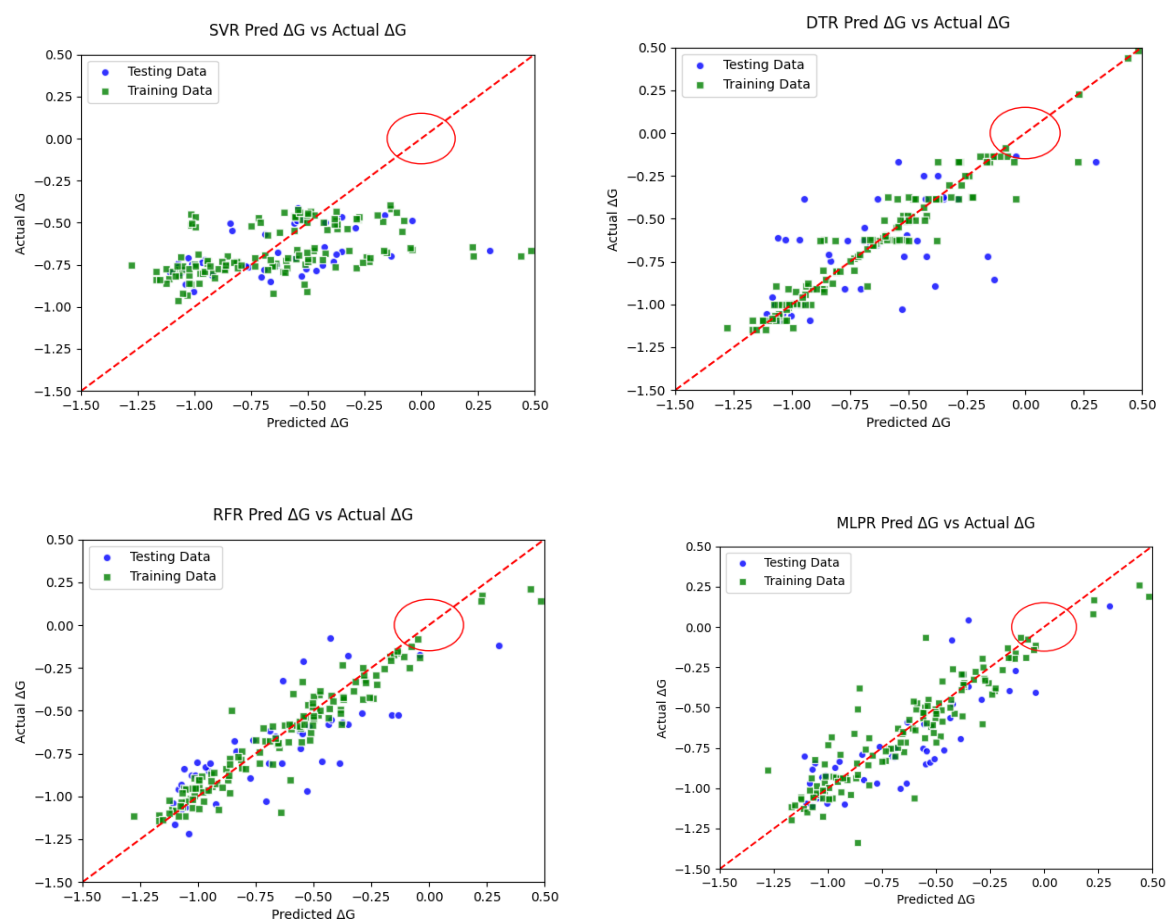


Figure 3. Graph of Actual Delta G vs Predicted Delta G of training data and testing data using different models. Circle illustrates  $\pm 0.15$  eV of Delta G.

Table 2. Table of training and testing RMSE, and model generation and prediction time of all models.

	Training RMSE	Testing RMSE	Model Generation Time (s)	Model Prediction Time (s)
OLS	0.2504	0.2798	0.0124	0.0140
GPR	0.2516	0.2722	0.0697	0.0418
KRR	0.2665	0.2810	0.0146	0.0156
RidgeCV	0.2528	0.2727	0.0127	0.0130
SVR	0.3454	0.3469	0.0128	0.0196
DTR	0.0894	0.3304	0.0105	0.0129
RFR	<u>0.1011</u>	<u>0.2702</u>	<u>0.1107</u>	<u>0.1247</u>
MLPR	0.1357	0.2978	1.7014	1.6907

Table 3. Materials predicted by RFR.

Formula	Predicted E
RhB2	0.108538
Rh3B4	0.117085
Rh5B2	0.106781
RhC2	0.072489
Rh3C4	0.079534
Rh5C2	0.069535
RhN2	-0.09635
Rh3N4	-0.14667
NiB	-0.06545
NiB2	0.111525
Ni2B	-0.04902
Ni3B4	0.13049
Ni5B2	0.128719
NiC	-0.08873
NiC2	0.076198
Ni2C	-0.0723
Ni3C4	0.096403
Ni5C2	0.091658
NiN2	-0.06882
Ni2N	-0.1559
Ni3N4	-0.13008
CoB2	-0.09202
Co3B4	-0.0812
Co5B2	-0.0851
CoC2	-0.09768
Co3C4	-0.08792
Co5C2	-0.09439

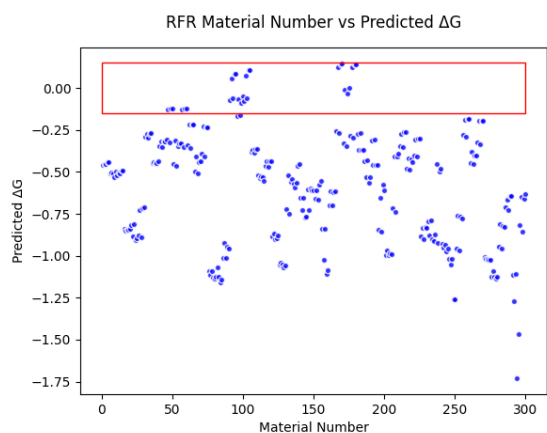


Figure 4. Graph of all 300 materials predicted, red square illustrates  $\pm 0.15$  eV of Delta G.

## Discussion

We found that Random Forest Regression (RFR) produced the lowest testing RMSE and was chosen to predict our 300 materials. 29 materials were within  $\pm 0.15$  eV of Delta G, this is roughly a 10-fold reduction in materials to be assessed with DFT. Models were not extensively optimized and used only one descriptor related to actual material composition, that being the M/X ratio.

The speed and relative ease of building and using machine learning models does seem to be a promising screening tool for HER catalyst discovery. Our results for some models are very similar in accuracy to other studies utilizing more advanced descriptors such as Bader charge transfer and this shows promise for using simpler descriptors such as in our study to accurately predict Gibbs free energy change for materials. Our results seem to indicate that RFR is the ideal supervised model for predicting HER activity.

## References

- (1) Wee, H.-M.; Yang, W.-H.; Chou, C.-W.; Padilan, M. V. Renewable Energy Supply Chains, Performance, Application Barriers, and Strategies for Further Development. *Renew. Sustain. Energy Rev.* **2012**, *16* (8), 5451–5465.
- (2) Han, X.; Lu, L.; Zheng, Y.; Feng, X.; Li, Z.; Li, J.; Ouyang, M. A Review on the Key Issues of the Lithium Ion Battery Degradation among the Whole Life Cycle. *eTransportation* **2019**, *1*, 100005.
- (3) Kaunda, R. B. Potential Environmental Impacts of Lithium Mining. *J. Energy Nat. Resour. Law* **2020**, *38* (3), 237–244.
- (4) Wanger, T. C. The Lithium Future—Resources, Recycling, and the Environment. *Conserv. Lett.* **2011**, *4* (3), 202–206.
- (5) Stamp, A.; Lang, D. J.; Wäger, P. A. Environmental Impacts of a Transition toward E-Mobility: The Present and Future Role of Lithium Carbonate Production. *J. Clean. Prod.* **2012**, *23* (1), 104–112.
- (6) Jain, I. P. Hydrogen the Fuel for 21st Century. *Int. J. Hydrog. Energy* **2009**, *34* (17), 7368–7378.

- (7) Veziroğlu, T. N.; Barbir, F. Hydrogen: The Wonder Fuel. *Int. J. Hydrog. Energy* **1992**, *17* (6), 391–404.
- (8) Haeseldonckx, D.; D'haeseleer, W. The Use of the Natural-Gas Pipeline Infrastructure for Hydrogen Transport in a Changing Market Structure. *Int. J. Hydrog. Energy* **2007**, *32* (10), 1381–1386.
- (9) Foh, S.; Novil, M.; Rockar, E.; Randolph, P. *Underground Hydrogen Storage. Final Report. [Salt Caverns, Excavated Caverns, Aquifers and Depleted Fields]*; BNL-51275; Brookhaven National Lab., Upton, NY (USA), 1979.
- (10) Ozarslan, A. Large-Scale Hydrogen Energy Storage in Salt Caverns. *Int. J. Hydrog. Energy* **2012**, *37* (19), 14265–14277.
- (11) Lord, A. S. *Overview of Geologic Storage of Natural Gas with an Emphasis on Assessing the Feasibility of Storing Hydrogen.*; SAND2009-5878; Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA (United States), 2009.
- (12) Züttel, A. Hydrogen Storage Methods. *Naturwissenschaften* **2004**, *91* (4), 157–172.
- (13) Barelli, L.; Bidini, G.; Gallorini, F.; Servili, S. Hydrogen Production through Sorption-Enhanced Steam Methane Reforming and Membrane Technology: A Review. *Energy* **2008**, *33* (4), 554–570.
- (14) Simpson, A. P.; Lutz, A. E. Exergy Analysis of Hydrogen Production via Steam Methane Reforming. *Int. J. Hydrog. Energy* **2007**, *32* (18), 4811–4820.
- (15) Kreuter, W.; Hofmann, H. Electrolysis: The Important Energy Transformer in a World of Sustainable Energy. *Int. J. Hydrog. Energy* **1998**, *23* (8), 661–666.
- (16) Zhu, J.; Hu, L.; Zhao, P.; Lee, L. Y. S.; Wong, K.-Y. Recent Advances in Electrocatalytic Hydrogen Evolution Using Nanoparticles. *Chem. Rev.* **2020**, *120* (2), 851–918.
- (17) Zhao, G.; Rui, K.; Dou, S. X.; Sun, W. Heterostructures for Electrochemical Hydrogen Evolution Reaction: A Review. *Adv. Funct. Mater.* **2018**, *28* (43), 1803291.
- (18) Sheng, W.; Gasteiger, H. A.; Shao-Horn, Y. Hydrogen Oxidation and Evolution Reaction Kinetics on Platinum: Acid vs Alkaline Electrolytes. *J. Electrochem. Soc.* **2010**, *157* (11), B1529.
- (19) Schuldiner, S. Hydrogen Overvoltage on Bright Platinum: II. pH and Salt Effects in Acid, Neutral, and Alkaline Solutions. *J. Electrochem. Soc.* **1954**, *101* (8), 426.
- (20) Conway, B. E.; Bai, L. Determination of Adsorption of OPD H Species in the Cathodic Hydrogen Evolution Reaction at Pt in Relation to Electrocatalysis. *J. Electroanal. Chem. Interfacial Electrochem.* **1986**, *198* (1), 149–175.
- (21) Zheng, Y.; Jiao, Y.; Vasileff, A.; Qiao, S.-Z. The Hydrogen Evolution Reaction in Alkaline Solution: From Theory, Single Crystal Models, to Practical Electrocatalysts. *Angew. Chem. Int. Ed.* **2018**, *57* (26), 7568–7579.
- (22) Lukowski, M. A.; Daniel, A. S.; Meng, F.; Forticaux, A.; Li, L.; Jin, S. Enhanced Hydrogen Evolution Catalysis from Chemically Exfoliated Metallic MoS<sub>2</sub> Nanosheets. *J. Am. Chem. Soc.* **2013**, *135* (28), 10274–10277.
- (23) Gao, G.; O'Mullane, A. P.; Du, A. 2D MXenes: A New Family of Promising Catalysts for the Hydrogen Evolution Reaction. *ACS Catal.* **2017**, *7* (1), 494–500.
- (24) Sun, X.; Zheng, J.; Gao, Y.; Qiu, C.; Yan, Y.; Yao, Z.; Deng, S.; Wang, J. Machine-Learning-Accelerated Screening of Hydrogen Evolution Catalysts in MBenes Materials. *Appl. Surf. Sci.* **2020**, *526*, 146522.
- (25) Li, Z.; Wang, S.; Chin, W. S.; Achenie, L. E.; Xin, H. High-Throughput Screening of

- Bimetallic Catalysts Enabled by Machine Learning. *J. Mater. Chem. A* **2017**, 5 (46), 24131–24138.
- (26) Ghahramani, Z. Unsupervised Learning. In *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures*; Bousquet, O., von Luxburg, U., Rätsch, G., Eds.; Lecture Notes in Computer Science; Springer: Berlin, Heidelberg, 2004; pp 72–112.
- (27) MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. *Proc. Fifth Berkeley Symp. Math. Stat. Probab. Vol. 1 Stat.* **1967**, 5.1, 281–298.