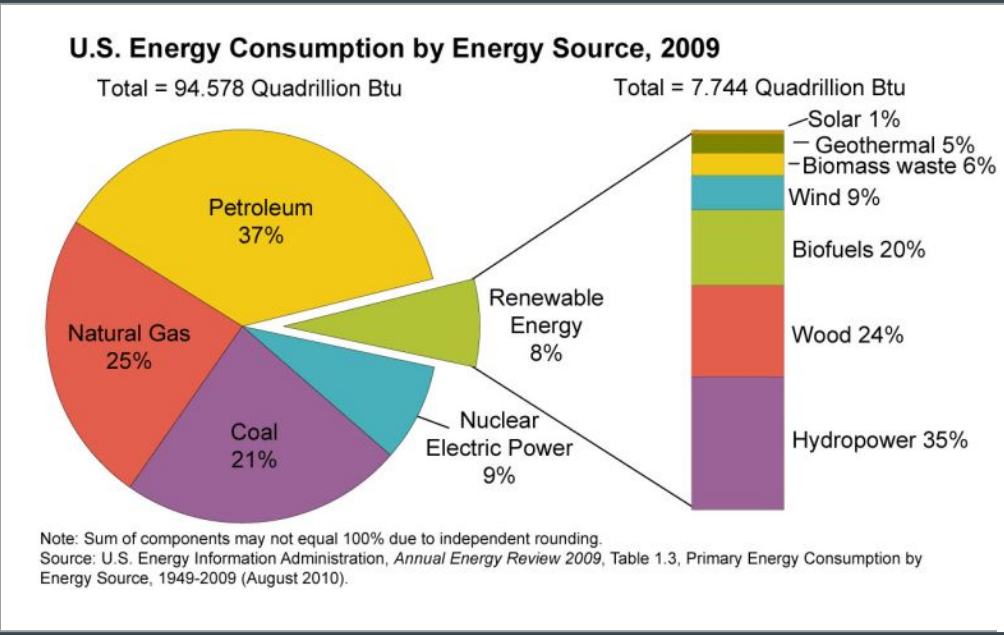
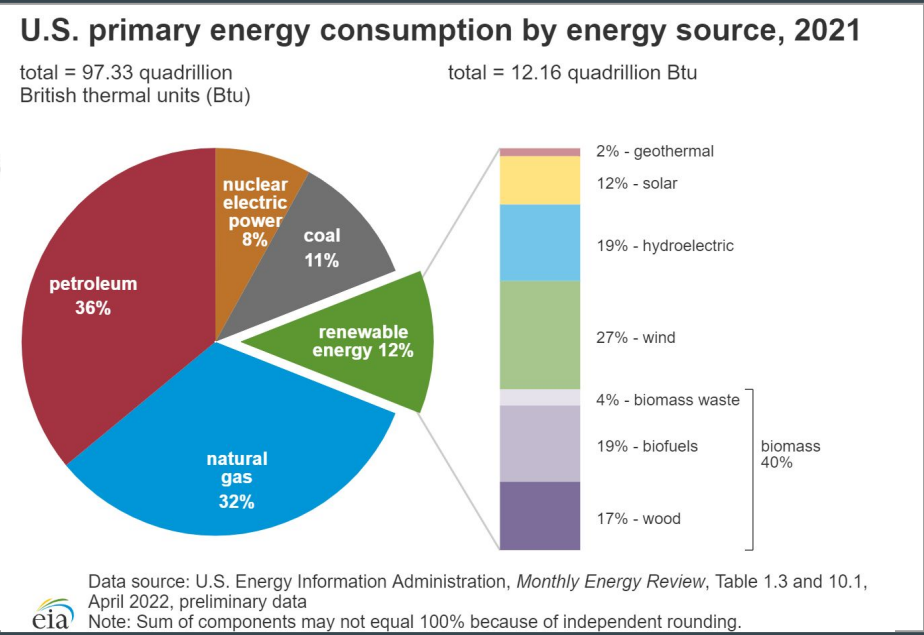
Screening of Highly Active Hydrogen Evolution Reaction Catalysts: Comparative Analysis of Typical Machine Learning Methods

Warren Wilson College
Jordan Mershimer
4/14/2023

The Energy Problem





- Renewable Energy: naturally replenishing but **flow-limited**
- Non-renewable Energy: Will not be replenished for thousands or even millions of years

Energy Storage approaches

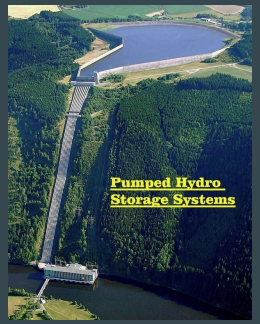
Batteries

- Degrades over time
- Some require rare-earth minerals to manufacture
- Manufacturing process could be detrimental to the environment

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.

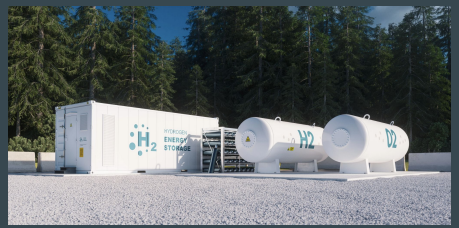
Pumped Storage Hydropower

- Costly to build
- Sizable physical footprint
- Not applicable to all environments



Hydrogen Gas

- Can use existing fossil fuel infrastructure
- Produces water when combusted
- Can be stored as liquid, solid, or gas.

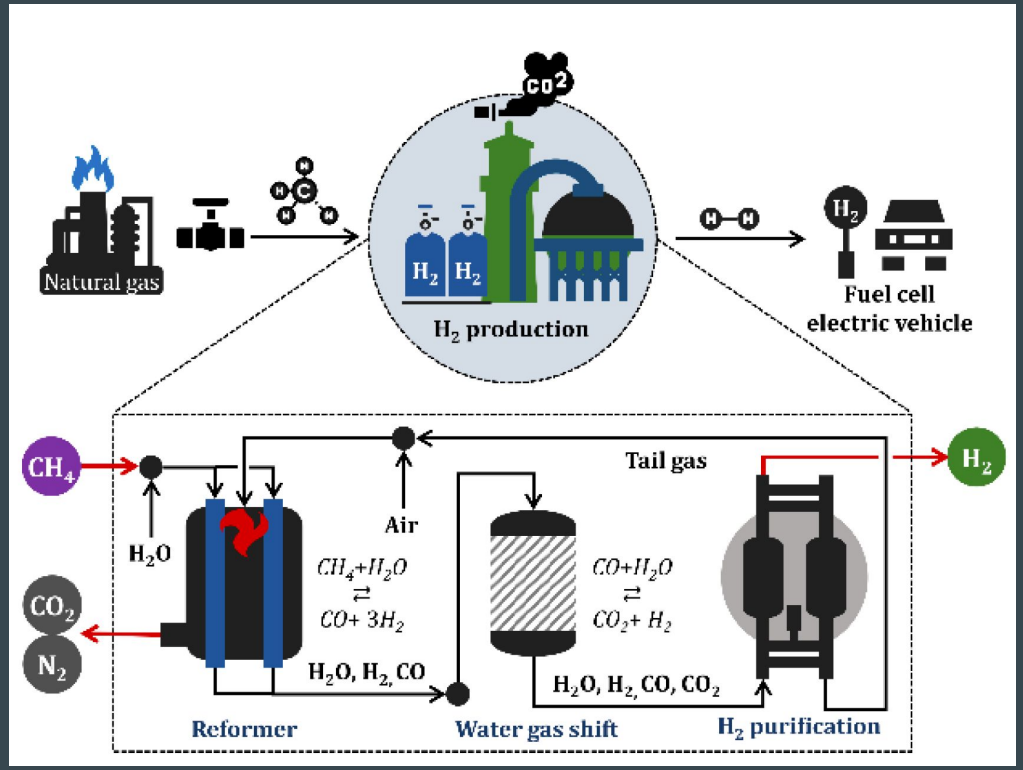


Züttel, A. Hydrogen Storage Methods. *Naturwissenschaften* **2004**, *91* (4), 157–172.

Hydrogen gas production

Steam Methane Reforming

- 80-85% of current hydrogen gas produced is by steam methane reforming
- Relies on fossil fuels
- Produces CO_2



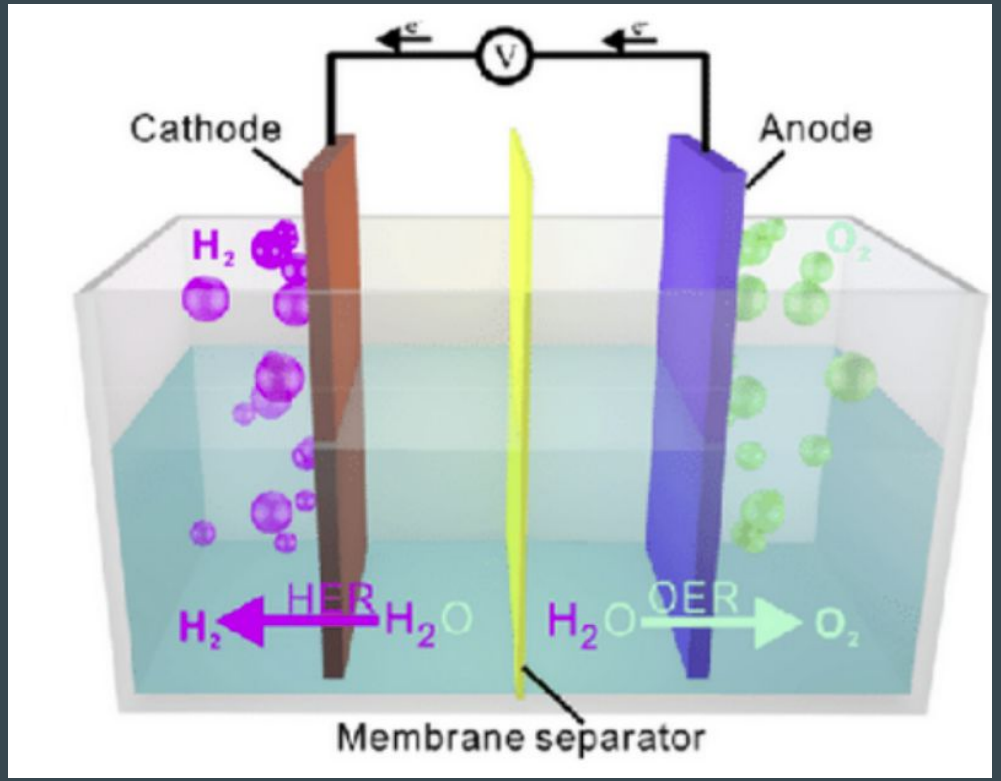
(13) Barelli, L.; Bidini, G.; Gallorini, F.; Servili, S. Hydrogen Production through Sorption-Enhanced Steam Methane Reforming and Membrane Technology: A Review. *Energy* **2008**, *33* (4), 554–570.

(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.

Hydrogen gas production

Water Electrolysis

- Water can be split into hydrogen gas and oxygen through Water Electrolysis.
- This reaction requires an overpotential of energy to occur, reducing the efficiency.
- The most efficient catalysts are composed of noble-metals elements like Pt, Ru, ...



(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.

Reprinted from ref 21. 2020 Zhu, J.

Indicator of a good catalyst

Gibbs Free Energy

- The interactions between the catalyst and Hydrogen should be "just right".
- Hydrogen adsorption free energy (ΔG_H) close to zero.

Active Sites

- Large number of sites where HER can occur on the surface of the material.

Stable

- Stable in wide range of pH conditions.
- Easy to work with and apply to cathodes.

(22) Lukowski, M. A.; Daniel, A. S.; Meng, F.; Forticaux, A.; Li, L.; Jin, S. Enhanced Hydrogen Evolution Catalysis from Chemically Exfoliated Metallic MoS2 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.
Mou, J.; Gao, Y.; Wang, J.; Ma, J.; Ren, H. Hydrogen Evolution Reaction Activity Related to the Facet-Dependent Electrocatalytic Performance of NiCoP from First Principles. *RSC Advances* 2019, *9* (21), 11755–11761. https://doi.org/10.1039/C9RA01560D.

Indicator of a good catalyst

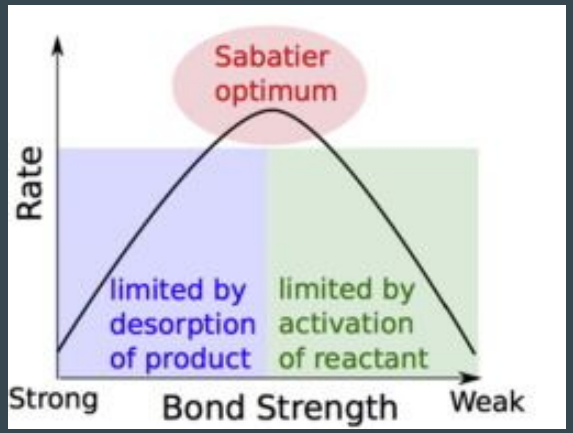
Sabatier principle:

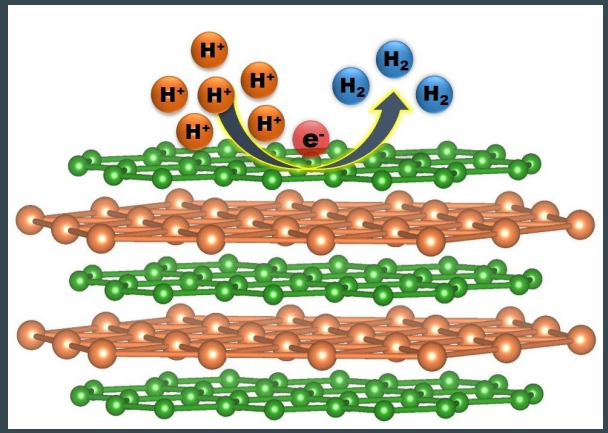
The interactions between the catalyst and Hydrogen should be "just right"; that is, neither too strong nor too weak.

Hydrogen adsorption free energy (ΔG_H) close to zero.

$$M + \tfrac{1}{2} H_2 \rightarrow M^*\text{-}H$$

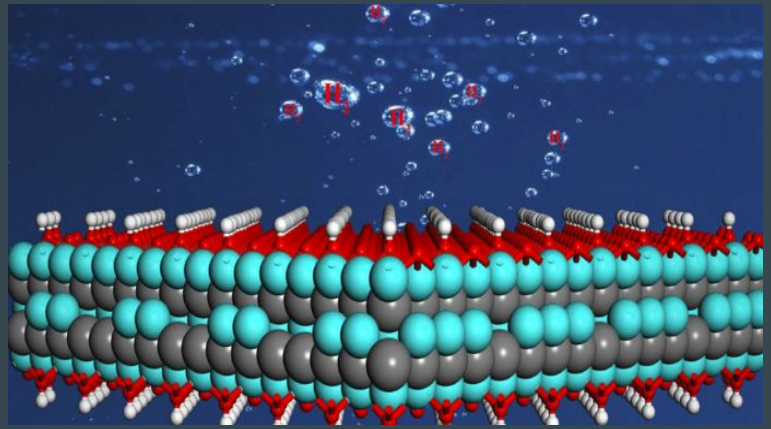
Surface Surface-H”





Mou, J.; Gao, Y.; Wang, J.; Ma, J.; Ren, H. Hydrogen Evolution Reaction Activity Related to the Facet-Dependent Electrocatalytic Performance of NiCoP from First Principles. *RSC Advances* **2019**, *9* (21), 11755–11761. https://doi.org/10.1039/C9RA01560D.

Catalyst Progress



2D materials:

2-dimensional materials offer unique advantages as catalysts.

- Can be uniformly deposited on a cathode
- Large surface area

Current 2D HER catalysts offer some disadvantages..

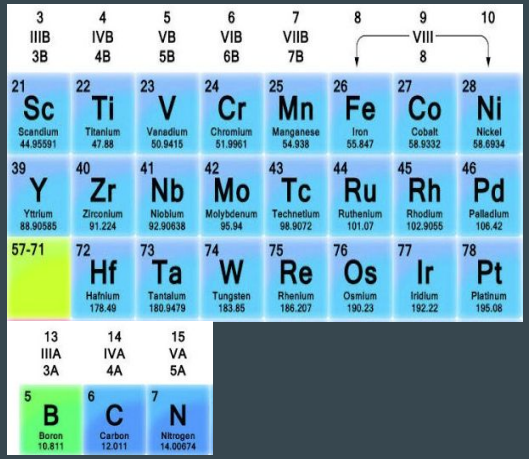
- Low number of active sites.
- Poor charge-transfer performance
- Unstable under ambient conditions

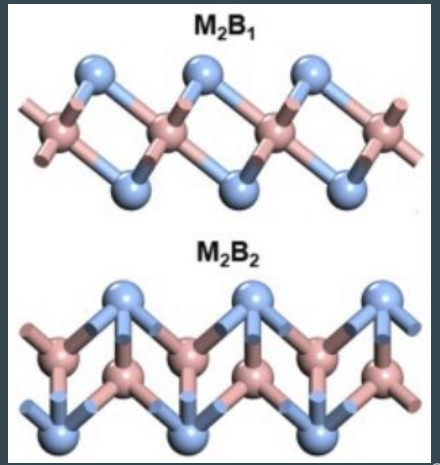
MXenes & MBenes:

2-dimensional materials composed of early transitional metal (M) from 3B, 4B, and 5B on the periodic table and Boron (B) or Nitrogen, Carbon (X).

Details:

- Inexpensive composition
- Not thoroughly explored





(22) Lukowski, M. A.; Daniel, A. S.; Meng, F.; Forticaux, A.; Li, L.; Jin, S. Enhanced Hydrogen Evolution Catalysis from Chemically Exfoliated Metallic MoS2 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.

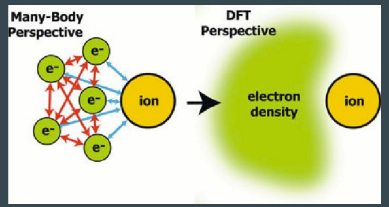
Catalyst discovery

Experimentally

- Synthesizing a single material can take a few days
- Requires lab time, materials, and active work.
- Results are completely accurate

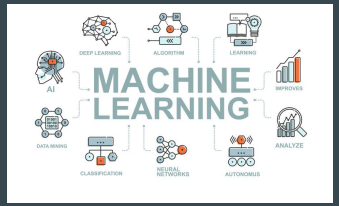
Density Functional Theory

- DFT is a computational method for modeling electronic structures of many-body systems.
- One result can take a few days or more
- Usually very accurate



Machine Learning

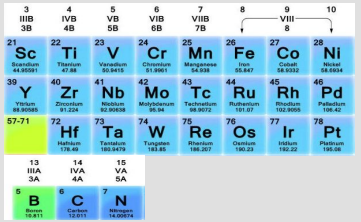
- Machine learning is the usage of computer algorithms to predict data.
- Depending on sample size, a model can be generated within seconds.
- Can be quite a complicated process



Pratima Menyangbo

Project objective:

Compare typical machine learning methods for Hydrogen Evolution Reaction (HER) catalyst screening



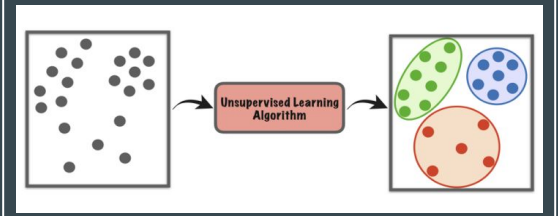
Focus: MXenes, MBenes

Early Transitional Metals (M). Boron (B), Carbon, Nitrogen (X). Ratios: 1:1, 1:2, 2:1, 3:4, 5:2.

Types of machine learning

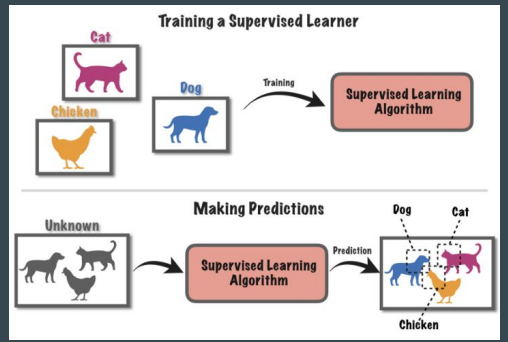
Unsupervised

- No labels
- No feedback
- Find hidden structures or associations in data



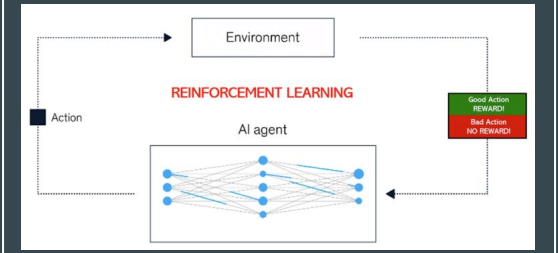
Supervised

- Labeled data
- Direct feedback
- Predict outcomes



Reinforcement

- Decision Process
- Reward System
- Learn series of actions to predict outcome



(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.

Methods

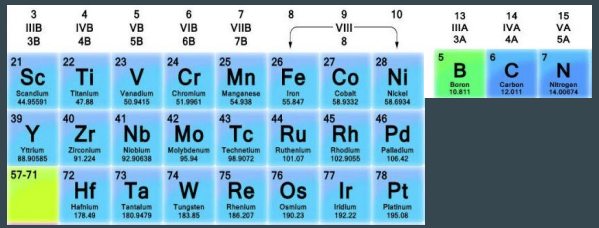
Models: Kernel Ridge, Decision Tree, Gaussian Process, Ordinary Least Squares, Multilayer Perceptron, Support Vector Machines, Random forest, Kernel Ridge Cross Validation.

Comparison: Root Mean Squared Error, Model Generation and Prediction Time.

Model Generation (published data.[24]): M_nB_m, M_nX_m.
n:m = 2:1, 2:2, 2:3, 4:3, 3:2.
Various known catalysts.

Materials Predicted: M_nB_m, M_nX_m.
n:m = 1:1, 1:2, 2:1, 3:4, 5:2.
Training:Testing Split = 75%:25%.
300 Materials were predicted

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	



(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.

Methods

Check out the code, modify it, use it.
https://github.com/Jmersh/NSURS-HER-ML

	Metal Symbol	M Number	X Symbol	X Number	Metal Sym Number	Metal Group	Metal Row	Metal Mass	Metal Density	Metal Inter-atomic Distance	Metal Covalent Radius	Metal First Ionization Energy	Metal Electron Affinity	X Sym Number	X Group	X Row	X Mass	X Density	X Inter-atomic Distance	X Covalent Radius	X First Ionization Energy	X Electron Affinity
IrC	Ir	1	C	1	77	9	6	192.217	22.42	2.423641628	1.41	8.96702	1.5643615	77	14	2	12.0107	2.2	2.085122401	0.76	11.260288	1.262113612
IrC2	Ir	1	C	2	77	9	6	192.217	22.42	2.423641628	1.41	8.96702	1.5643615	77	14	2	12.0107	2.2	2.085122401	0.76	11.260288	1.262113612
Ir2C	Ir	2	C	1	77	9	6	192.217	22.42	2.423641628	1.41	8.96702	1.5643615	77	14	2	12.0107	2.2	2.085122401	0.76	11.260288	1.262113612
Ir3C4	Ir	3	C	4	77	9	6	192.217	22.42	2.423641628	1.41	8.96702	1.5643615	77	14	2	12.0107	2.2	2.085122401	0.76	11.260288	1.262113612
Ir5C2	Ir	5	C	2	77	9	6	192.217	22.42	2.423641628	1.41	8.96702	1.5643615	77	14	2	12.0107	2.2	2.085122401	0.76	11.260288	1.262113612

```python
for M_MatN2 in M_Mat:  # Two for loops are nested together to iterate through each combination of M and X
    # materials, could use simpler method.
    for X_MatN2 in X_Mat:
        ps = pt.elements.symbol  # Declare ps as shorthand for longer function
        pyme = pymatgen.core.periodic_table.Element  # Pymatgen implementation
        MatGenTestCSVWrite.writerows([((M_MatN2 + X_MatN2), M_MatN2, "1", *X_MatN2, "1",
                                       pyme(M_MatN2).number, pyme(M_MatN2).group, pyme(M_MatN2).row,
                                       ps(M_MatN2).mass, ps(M_MatN2).density, ps(M_MatN2).interatomic_distance,
                                       ps(M_MatN2).covalent_radius, pyme(M_MatN2).ionization_energy,
                                       pyme(M_MatN2).electron_affinity,
                                       pyme(M_MatN2).number, pyme(X_MatN2).group, pyme(X_MatN2).row,
                                       ps(X_MatN2).mass, ps(X_MatN2).density, ps(X_MatN2).interatomic_distance,
                                       ps(X_MatN2).covalent_radius, pyme(X_MatN2).ionization_energy,
                                       pyme(X_MatN2).electron_affinity)])
        # Above writes rows of each iteration of the M and X materials along with descriptors such as
        # mass etc.
```

```python
def plotdata(model):
    testdata = pd.read_csv('Models/' + model + '/TestArray.csv')
    traindata = pd.read_csv('Models/' + model + '/TrainArray.csv')
    fig, ax = plt.subplots()

    ax.scatter([testdata.iloc[:, 0]], [testdata.iloc[:, 1]], color='blue', marker='o', edgecolors='white', alpha=0.8)
    ax.scatter([traindata.iloc[:, 0]], [traindata.iloc[:, 1]], color='green', marker='s', edgecolors='white', alpha=0.8)
    ax.legend(['Testing Data', 'Training Data'])
    plt.axis([-1.5, 0.5, -1.5, 0.5])
    plt.title(model + " Pred E vs Actual E", pad=15)
    plt.xlabel("Predicted ΔG")
    plt.ylabel("Actual ΔG")
    line = mlines.Line2D([0, 1], [0, 1], color='red', linestyle='dashed')
    transform = ax.transAxes
    line.set_transform(transform)
    ax.add_line(line)
    circlehighlight = plt.Circle((0, 0), 0.15, color='r', fill=False)
    plt.gca().add_patch(circlehighlight)
    plt.savefig('Models/' + model + '/TrainVSTest.png')
    plt.show()
```

```python
# RFR
t.start()
rfr = RandomForestRegressor(max_depth=20)
GenModel(rfr, 1).ReturnArray("RFR")
t.stop()

t.start()
GenModel(rfr, 1, p=1).ReturnArray("RFR", p=1)
t.stop()
```

```python
plotdata('DTR')
plotdata('GPR')
plotdata('KRR')
plotdata('MLPR')
plotdata('OLS')
plotdata('RFR')
plotdata('RidgeCV')
plotdata('SVR')
```

- Kmeans Clustering Intro.py
- KmeansMatRender.py
- LearningModelGen.py
- LearningModelGenBeta.py
- MatGen.py
- MatGenOutput.csv
- MatGenRender.py
- ModelRender.py
- PyPlot Intro.py
- README.md
- TrainData.csv
- TrainDataGen.py
- TrainDataOutput.csv
- TrainingData.csv

- Models
  - DTR
    - MatGenArray.csv
    - RMSEArray.csv
    - RMSEBar.png
    - TestArray.csv
    - TrainArray.csv
    - TrainVSTest.png
  - GPR
  - KRR
  - MLPR
  - OLS
  - RFR
  - RidgeCV
  - SVR

Comparison

We found that Random Forest Regression (RFR) produced the lowest testing RMSE and was chosen to predict our 300 materials.

Models were trained on published training data[24].

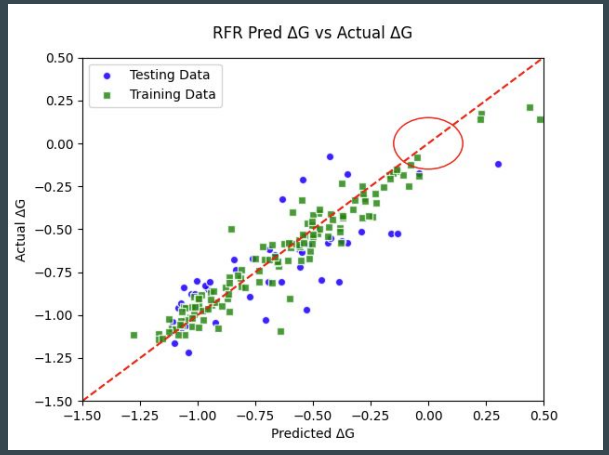
Models are not completely optimized.

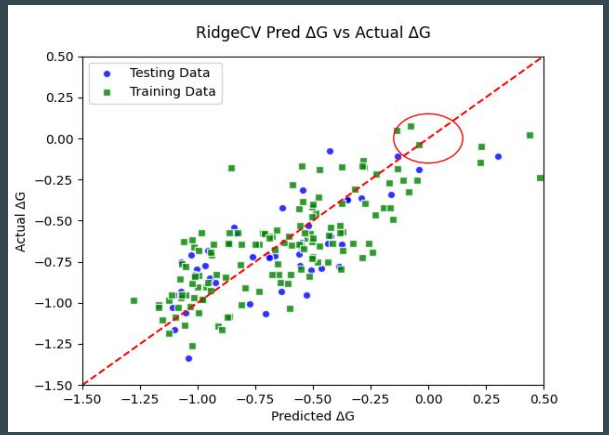
	Training RMSE	Testing RMSE	Model Generation Time (s)	Model Prediction Time (s)
OLS	0.2504	0.2798	0.0124	0.0140
GPR	0.2516	0.2722	0.0697	0.0418
KRR	0.2665	0.2810	0.0146	0.0156
RidgeCV	0.2528	0.2727	0.0127	0.0130
SVR	0.3454	0.3469	0.0128	0.0196
DTR	0.0894	0.3304	0.0105	0.0129
RFR	**0.1011**	**0.2702**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

Lower RMSE is better. RMSE Results are averaged over 10 different training/testing split randomizations. Predicted 300 materials.

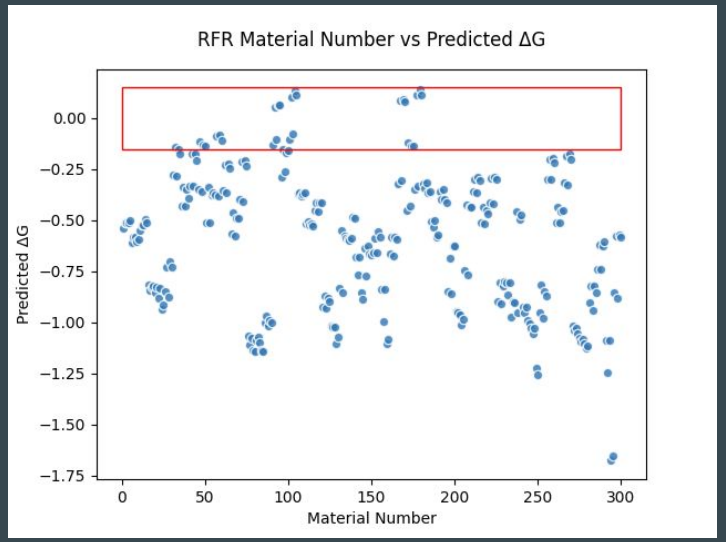
(24) Sun, X.; Zheng, J.; Gao, Y.; Qiu, C.; Yan, Y.; Yao, Z.; Deng, S.; Wang, J. Machine-Learning-Accelerated Screening of Hydrogen Evolution Catalysts in MBenes Materials. *Appl. Surf. Sci.* **2020**, *526*, 146522.

Model Prediction Results





Random Forest Regression (RFR)



Number	Predicted Gibbs Free Energy	Name	Screen
32	-0.13946	RuC2	Yes
34	-0.15172	Ru3C4	Yes
47	-0.11625	CoC2	Yes
49	-0.13134	Co3C4	Yes
91	-0.1281	NiC	Yes
92	0.053562	NiC2	Yes

29 materials were within ±0.16 eV of 0.
Table represents some data. 10-fold reduction.

Summary

- 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.

- 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.

- Future research could include experimental verification of our 29 materials or further research and optimization of our machine learning approach.

Project Team

Jordan Mershimer

Me

Yuemei Zhang, Ph.D.

Head Advisor

Langdon Martin, Ph.D.

Advising Committee

Kim Borges, Ph.D.

Advising Committee

Acknowledgement

Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

The Matplotlib development team

Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent Chevrier, Kristin A. Persson, Gerbrand Ceder. Python Materials Genomics (pymatgen) : A Robust, Open-Source Python Library for Materials Analysis. Computational Materials Science, 2013, 68, 314–319. doi:10.1016/j.commatsci.2012.10.028

The pandas development team

Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. Nature 585, 357–362 (2020).

San Diego Supercomputer Center

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) Veziro»lu, 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 MoS2 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.