# Machine Learning-based Exploration of the Effect of Promoters on Transition Metal Carbide Catalysts for Hydrogen Evolution Reaction

M.Sc. Thesis

By NAMAN CHAUDHARY



# DEPARTMENT OF CHEMISTRY INDIAN INSTITUTE OF TECHNOLOGY INDORE MAY 2023

# Machine Learning-based Exploration of the Effect of Promoters on Transition Metal Carbide Catalysts for Hydrogen Evolution Reaction

### A THESIS

Submitted in partial fulfilment of the requirements for the award of the degree

of

### **MASTER OF SCIENCE**

by

### NAMAN CHAUDHARY



# DEPARTMENT OF CHEMISTRY INDIAN INSTITUTE OF TECHNOLOGY INDORE MAY 2023



# INDIAN INSTITUTE OF TECHNOLOGY INDORE

### **CANDIDATE'S DECLARATION**

I hereby certify that the work which is being presented in the thesis entitled Machine Learning-based Exploration of the Effect of Promoters on Transition Metal Carbide Catalysts for Hydrogen Evolution Reaction. In the fulfilment of the requirements for the award of the degree of MASTER OF SCIENCE (RESEARCH) and submitted in the DEPARTMENT OF CHEMISTRY, Indian Institute of Technology Indore, is an authentic record of my own work carried out during the time period from August 2022 to May 2023 under the supervision of Prof. Biswarup Pathak, Department of Chemistry, Indian Institute of Technology Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Signature

(Naman Chaudhary)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of the Supervisor with date

(Prof. Biswarup Pathak)

Naman Chaudhary has successfully given his M.Sc. (Research) OralExamination held on Date of

Signature(s) of Supervisor(s) of MSc thesis Prof. Biswarup Pathak

Signature of PSPC Member 24/5

Dr. Tushar Kanti Mukherjee

Date:

Convener, DPGC

jekz 24/05/2023

Date:

Span len fr

Dr. Dipak Kumar Roy Date: 24.05.2023

#### Acknowledgments

I am extremely grateful to Prof. Biswarup Pathak, my thesis supervisor and a Professor in the Department of Chemistry at the Indian Institute of Technology Indore, for granting me the opportunity to undertake my research project within his esteemed research group. His continuous guidance, timely support, and motivation have played a crucial role in the successful completion of my M.Sc. project. His enthusiasm and dedication have been a constant source of inspiration for me. I would also like to express my heartfelt appreciation to the members of my thesis committee, Dr. Tushar Kanti Mukherjee and Dr. Dipak Kumar Roy, for their valuable suggestions and kind support throughout the research process. Furthermore, I am thankful to Prof. Suhas S. Joshi, the Director of the Indian Institute of Technology Indore, for his encouragement and for providing the necessary facilities at the institute. I am indebted to the Indian Institute of Technology Indore for providing me with laboratory and computational facilities during my research project. I would like to acknowledge the support and motivation I received from the Head of the Department of Chemistry, Dr. Tushar Kanti Mukherjee as well as from Prof. Rajneesh Misra, Prof. Suman Mukhopadhyay, Prof. Apurba K. Das, Prof. Sampak Samanta, Dr. Anjan Chakraborty, Dr. Shaikh M. Mobin, Dr. Satya S. Bulusu, Dr. Chelvam Venkatesh, Dr. Amrendra Kumar Singh, Dr. Abhinav Raghuvanshi, Dr. Dipak Kumar Roy, Dr. Selva kumar Sermadurai, and Dr. Umesh A. Kshirsagar throughout my M.Sc. course. I am deeply appreciative of the exceptional professional support provided by the Chemistry office staff, namely Mr. Manish Kushwaha, Mr. Parthiban, Mrs. Vinita Kothari, and Mr. Rameshwar Dauhare. I would also like to extend my special thanks to my project mentor, Mr. Diptendu Roy, and Mr. Amitabh Das for their constant help and guidance throughout the project. I feel fortunate to have had the privilege of working alongside senior lab members Dr. Mohan Tiwari, Mr. Sandeep Das, Mr. Arunendu Das, Mr. Surya Shekhar Manna, Mr. Milan

Kumar Jena, Mrs. Nishchal Bharadwaj, Ms. Eti Mahal, Ms. Sneha Mittal, Mr. Souvik Manna, Ms. Harpriya Minhas, and Mr. Rahul Kumar Sharma. Their assistance in various aspects, friendly discussions, shared lunch breaks, memorable treats, and enjoyable moments outside the lab have been truly invaluable. I would also like to express my gratitude to my friends Shiwani, Sacchu, Meena, Nitish, Aado, Preru, Tussu, and Rathi, who have been an integral part of my M.Sc. journey. Lastly, I extend my sincere thanks to my parents, Mr Ravi Kant and Mrs. Neeta Rani, my brother Shivam Chaudhary, and all my family members for their unwavering trust and support throughout my academic pursuits

(Naman Chaudhary)

# Dedication

This thesis is dedicated to my parents and friends.

#### Abstract

The development of cost-effective electrocatalysts to substitute expensive platinum group metals (PGMs) in catalyzing the hydrogen evolution reaction (HER) holds immense promise for advancing sustainable energy solutions. In recent years, machine learning (ML) techniques have opened up new possibilities for smart screening and prediction of efficient heterogeneous catalysts. ML methods offer powerful tools that can analyze large datasets, extract meaningful patterns and correlations, and make accurate predictions based on the learned information. We utilized combined density functional theory (DFT) with supervised machine learning methods to discover earth abundant, durable, low-cost, efficient, transition metal carbides (TMCs) based active heterogeneous catalyst for HER. The study utilizes a Kernel Ridge Regression (KRR) model that has been optimized to identify the catalyst that is most active towards HER. The study demonstrates that this approach enables the efficient screening of a vast array of catalysts, identifying the most active HER catalysts. Hence, our methodology offers an effective means to find heterogeneous catalysts suitable for diverse electrochemical reactions.

## **Table of Contents**

A. List of Figures	ix
<b>B.</b> List of Tables	xi
C. Acronyms	xiii

### **1. CHAPTER 1: Introduction**

1.1. Hydrogen Evolution Reaction (HER)	1
1.2. Choice of Catalysts	
1.3. Promoter Selection	
1.4. Artificial Intelligence (AI)	6
1.4.1. Machine Learning (ML)	7
1.4.1.1. Supervised Learning	8
1.4.1.2. Features	9
1.4.1.3. ML models	9
1.4.1.4. Testing and Validation	12
1.4.1.5. Hyperparameter tunning	13
2. CHAPTER 2: Computational Methods and	
Approximation	
2.1. Schrödinger Equation	14
2.2. Born-Oppenheimer Approximation	14
2.3. Hartree-Fock (HF) Approximation	15
2.4. Density Functional Theory (DFT)	16
2.4.1. The Hohenberg-Kohn Theorems	17

2.4.2. Kohn-Sham Formulation	18
2.4.3. Exchange-correlation Functional	19
2.4.4. Local Density Approximation (LDA)	19
2.4.5. Generalized Gradient Approximation	20
2.5. Basis Sets	20
2.5.1 Projector Augmented Wave (PAW) Method	21
2.6. GPAW and ASE	22
2.7. Computational Details	22

### 3. CHAPTER 3: Results and Discussions

3.1. Catalyst Modelling	25
3.2. Feature Selection	28
3.3. ML Model Selection	35
3.4. Model Optimization	36
3.5. Prediction Evaluation	36
3.6. Model Stability	41
3.7. Catalysts screening	42
4. CHAPTER 4: Conclusions 53	

### **5. REFERENCES**

55

### **List of Figures**

#### CHAPTER 1

Figure 1.1	Systematic representation of HER in acidic and	2
	alkaline medium	

Figure 1.2 Schematic representation of AI, ML and DL domain. 7

Figure 1.3Different types of methods in machine learning8

#### **CHAPTER 2**

- Figure 2.1 1 The 4×4×4 periodic slab of Hafnium carbide with 23 (100) surface. a) Top view and b) side view. The crossed spheres of the lower two layer signifies the fixed atoms, and uncrossed spheres represents the relaxed atoms for the adsorption energy calculations. Cyan colour represents Hf atoms and grey colour represents carbon atoms.
- Figure 2.2 The 4×4×4 periodic slab of Hafnium carbide with 23 (111) surface in which metal is present on the top layer and followed by carbon layer. a) Top view and b) side view. The crossed spheres of the lower two layer signifies the fixed atoms, and uncrossed spheres represents the relaxed atoms for the adsorption energy calculations. Cyan colour represents Hf atoms and grey colour represents carbon atoms.

Figure 2.3 The 4×4×4 periodic slab of Hafnium carbide with 24 (111) surface in which carbon is present on the top layer and followed by metal layer. a) Top view and b) side view. The crossed spheres of the lower two layer signifies the fixed atoms, and uncrossed spheres represents the relaxed atoms for the adsorption energy calculations. Cyan colour represents Hf atoms and grey colour represents carbon atoms.

#### **CHAPTER 3**

- Figure 3.1 (a) and (b) are the Top and side view of (100) 26 transition metal carbides surface, where cyan colour represents Hf and grey colour represent carbon. (c) represents, adsorption of H atom in a hollow site four-fold hollow (ffh) present in (100) surface and (d) represents, adsorption of promoters near the adsorbed hydrogen in the hollow site of metal carbide (100) surface. Here, cross sign indicates fixed atoms.
- Figure 3.2 (a) and (b) are the Top and side view of (111) 27 transition metal carbides surface, when metal is present on the top and cyan colour represents Hf and grey colour represent carbon. (c) and (d) represents, adsorption of H atom in fcc and hcp hollow site present in (111) surface, whereas (d) and (e) represents, adsorption of promoters near the adsorbed hydrogen in the fcc and hcp hollow sites of metal carbide (111) surface. Here, cross sign indicates fixed atoms.

- Figure 3.3 (a) and (b) are the Top and side view of (111) 27 transition metal carbides surface, when carbon is present on the top and cyan colour represents Hf and grey colour represent carbon. (c) and (d) represents, adsorption of H atom in fcc and hcp hollow site present in (111) surface, whereas (d) and (e) represents, adsorption of promoters near the adsorbed hydrogen in the fcc and hcp hollow sites of metal carbide (111) surface. Here, cross sign indicates fixed atoms.
- Figure 3.4 Graphical representation of diiferent sets of input 28 features each set consist of different elemental features of carbon, metal and promoters
- Figure 3.5 Representation of the feature importance of datasets 32 containg (111) and (100) surfaces having hcp and ffh hollow sites by considering 28 input Features of carbon, metal and promoters
- Figure 3.6 Representation of the feature importance of datasets 33 containg (111) and (100) surfaces having fcc and ffh hollow sites by considering 70 input Features of carbon, metal and promoters
- Figure 3.7Plot of predicted adsorption energies ( $\Delta E_{pred}$ ) versus38DFT calculated adsorption energies ( $\Delta E_{calc}$ ) using<br/>optimized KRR model for the datasets containing

(111) surfaces with fcc hollow sites and 100 surfaces with ffh

- Figure 3.8Plot of predicted adsorption energies ( $\Delta E_{pred}$ ) versus38DFT calculated adsorption energies ( $\Delta E_{calc}$ ) using<br/>optimized KRR model for the datasets containing<br/>(111) surface with hcp hollow sites and 100 surfaces<br/>with ffh
- Figure 3.9 Representation of MAE scores for datasets (111) and 41 (100) surfaces having hcp and ffh hollow sites, at 5 fold cross validation
- Figure 3.10 Representation of MAE scores for datasets (111) and 42 (100) surfaces having fcc and ffh hollow sites, at 5 fold cross validation

## **List of Tables**

### **CHAPTER 1**

#### **CHAPTER 2**

#### **CHAPTER 3**

- **Table 3.1**List of all the selected input features**30-31**
- Table 3.2Optimization of hyperparameters with their 39<br/>corresponding RMSE and MAE values for datasets<br/>containing (111) and (100) surfaces having fcc and<br/>ffh sites with the three considered regression<br/>models KRR, GBR and XGBR.

- **Table 3.3**Optimization of hyperparameters with their40corresponding RMSE and MAE values for datasetscontaining (111) and (100) surfaces having hcp andffh sites with the three regression models KRR,GBR and XGBR.
- **Table 3.4**Adsorption energies of different combination of**43-44**transition metal carbides and promoter for (100)surfaces containing four- fold hollow site
- **Table 3.5**Adsorption energies of different combination of<br/>transition metal carbides and promoter for (111)<br/>surfaces in which metal is on top containing fcc<br/>hollow site.
- **Table 3.6**Adsorption energies of different combination of47-48transition metal carbides and promoter for (111)surfaces in which metal is on top containing hcphollow site.
- Table 3.7Adsorption energies of different combination of<br/>transition metal carbides and promoter for (111)<br/>surfaces in which carbon is on top containing fcc<br/>hollow site.
- **Table 3.8**Adsorption energies of different combination of**51-52**transition metal carbides and promoter for (111)surfaces in which carbon is on top containing hcphollow site.

# Acronyms

AI	Artificial Intelligence
ASE	Atomistic Simulation Environment
DFT	Density Functional Theory
fcc	face-centered cubic
GGA	Generalized Gradient Approximation
GPAW	Grid-based Projector Augmented Wave
GBR	Gradient Boosting Regression
hcp	hexagonal close-packed
HER	Hydrogen Evolution Reaction
HF	Hartree-Fock
KRR	Kernel Ridge Regression
LDA	Local Density Approximation
ML	Machine Learning
PAW	Projector Augmented Wave
PBE	Perdew-Burke-Ernzerhof
PGM	Platinum Group Metal
RBF	Radial Basis Function
RMSE	Root Mean Square Error
SVM	Support Vector Machine
SVR	Support Vector Regression

- TMC Transition Metal Carbides
- XGBR Extreme Gradient Boosting Regression
- ZPE Zero-point Energy

#### Introduction

The world is experiencing an energy shortage currently.[1] Natural gas, fossil fuels, and other non-renewable sources have been unable to meet the country's energy needs because of their limited availability. [1,2] The continuing quest to realise this demand has found an excellent alternative in hydrogen because of its high energy density, minimal emissions of greenhouse gases or pollution, and renewable nature. [3,4] Today, methane, or natural gas, is reformatted using steam to produce the majority of the hydrogen used for industrial purposes. Due to natural gas's high availability and low cost, this production technology is favoured, however, natural gas can't be termed as a sustainable energy source, and the issue of CO<sub>2</sub> release is yet to be resolved. Although producing hydrogen through the interconversion of hydrogen and water offers a carbon-free, eco-friendly option [5], conducting the hydrogen evolution reaction (HER),  $(2H^+ + 2e^-)$ = H<sub>2</sub>) is still challenging without the aid of noble metal catalysts. For effective conduction of the HER process, a highly active catalyst is required. This catalyst is supposed to reduce the potential barrier of the catalytic process and also expedite the reaction.

#### **1.1. Hydrogen Evolution Reaction (HER)**

HER comprises two fundamental steps: the adsorption of hydrogen atoms onto the electrode surface and their subsequent detachment as a hydrogen molecule (H<sub>2</sub>). The kinetics of HER involve three well-known reactions: Volmer, Heyrovsky, and Tafel. The adsorption process is associated with the Volmer reaction, where hydrogen atoms are adsorbed onto the metal surface of the electrode due to the interaction between an electron and a proton in the surrounding medium. [6] On the other hand, both the Heyrovsky and Tafel reactions are primarily linked to desorption processes. In the Heyrovsky reaction, adsorbed two hydrogen atoms react with each other, resulting in the production of a hydrogen molecule. In the Tafel reaction, a hydrogen molecule has been formed from two adsorbed hydrogen atoms and emerged from the electrode surface without the involvement of an electron. It is possible for HER to occur in both acidic and alkaline environments, the only difference lying in the source of protons. In an acidic medium,  $H_3O^+$  serves as the proton source, while in an alkaline medium,  $H_2O$  provides the protons. Both acidic and alkaline environments support the HER, enabling the efficient generation of hydrogen gas.



**Figure 1.1** Systematic representation of HER in acidic and alkaline medium

#### **1.2.** Choice of Catalysts

According to Sabatier's principle [7], the efficiency of catalytic phenomena is maximized when there is an optimal interaction between the catalytic surface and intermediate hydrogen atoms (H\*), neither too weak nor too strong. This can be indicated by a low Gibbs free energy ( $\Delta G \sim 0$ ) or a moderate M-H bonding, which facilitates favorable H\*adsorption. Platinum (Pt) can be said to be an optimum choice. Pt is regarded as the most sustainable and effective catalyst, also because of its low overpotential and acceleration of reaction rates. Unfortunately, its limited availability and skyrocketing costs limit its usage. Two major limitations of Pt-based materials, aside from their high catalytic activity, need to be addressed for them to become viable for industrial use: (1) decreased catalyst loading; and (2) prevention against corrosion in harsh alkaline or acidic electrolytes. [8] As a result, the effective and efficient conduct of HER faces a lot of challenges. As this is happening, significant research is being done to find new non-noble metal electrocatalysts that are affordable, stable, and active to replace these priceless noble metals. In the field of electrocatalysis, a plethora of new material classes have recently emerged, including metal carbides, [9,10] nitrides, [11,12,13] and phosphides, as well as twodimensional (2D) materials such as layered double hydroxides (LDHs), metal oxides, MXenes, graphene, phosphorene, TM dichalcogenides (TMDs), and graphitic carbon nitride  $(g-C_3N_4)$ . [14] These materials have been extensively used in various electrocatalytic reactions and have demonstrated outstanding activity for HER. Although the previous studies have utilized TMCs for HER, our perspective is to investigate the combined effect of alkaline earth and alkali metals as promoters to further increase the efficiency, which still remains an unexplored area. There have been many thorough reviews of the many TMC properties. From earlier studies, we have observed many experimental findings for TMCs. [15,16,17,18] Moreover, TMCs of groups IV–VI early transition metals exhibit a variety of distinctive and fascinating catalytic characteristics. The catalytic capabilities of TMCs are the focus of numerous studies in the disciplines of surface science ever since the publication of seminal work by Levy and Boudart regarding the Pt-like properties of tungsten carbides. [19] According to Fujita and co-workers, as well as Li and co-workers, group IV-VI transition metal carbides have developed immense interest among

the scientific community to investigate their potential role in HER because they have the same properties as Pt group metals. [2] The merging of carbon atoms with the metal lattices results in the formation of TMCs. These carbide materials combine the distinguishing characteristics of three separate material classes— ionic crystals, covalent solids, and transition metals-to provide special physical and chemical properties. They exhibit the severe brittleness and hardness of covalent solids, the high melting point and basic crystal structures alike ionic crystals, and they resemble transition metals in terms of their electronic and magnetic properties. [20] The parent metal's d-band is changed by the creation of carbides, which results in catalytic characteristics that are distinct from the parent metals yet comparable to those of group VIII noble metals. The fact that the metal lattice extends and the metal-metal distance rises as a result of the development of interstitial compounds is one straightforward explanation. For instance, during carbide production, the lattice parameters of molybdenum and vanadium are respectively enlarged from 0.27 to 0.30 nm and from 0.26 to 0.42 nm. [21] Heine et al. have shown that the contraction of the metal d-band, is proportional to the inverse fifth power of the metalmetal distance, would occur when the metal-metal bond distance increased. [22,23] Despite the fact that there was a charge transfer from the parent metal to the carbon atoms, it is thought that a d-band contraction of this kind would produce a higher density of states (DOS) around the Fermi level than the parent metal.

An analysis of these compounds both theoretically and experimentally, supports the inclusion of carbon in the early transition metals' lattice. According to the calculations using DFT, the hybridization between the carbon s- and p-orbitals and metal d-orbitals lead to the broadening of the d-band structure which results properties similar to those of the d-band of Pt. These substances frequently exhibit greater catalytic efficiencies over their parent metals in terms of selectivity, activity, and toxin resistance. Therefore, in our work, we have taken 9 different TMCs from groups IV–

VI (TiC, ZrC, HfC, VC, NbC, TaC, CrC, MoC, WC) with (111) and (100) surfaces as an electrocatalyst for HER.

#### **1.3. Promoters Selection**

To increase the efficiency of the chemical reactions, promoters are added to the solid catalyst. These species do not exhibit any propensity to form conventional chemical connections with the electrode surface's active site. Basically, there are two types of interactions between the electrodes and adsorbates. First, is the direct bonding of adsorbates and electrodes, which entails electron transfer, chemisorption, and the release of the ion hydration shell. Second, relatively mild electrostatic metal-ion forces that do not directly interact with the adsorbate but may have an impact on the ion concentration near the electrode. These two kinds of bonds are connected with an extraordinarily wide variety of physicochemical activities, including as electron-charge transfer processes, ionic hydration, adsorption, and catalysis. [24] In order to increase HER efficiency, we use alkali and alkaline earth metals in our study as promoters that interact electrostatically with the TMCs. Nowadays, the promoter's role has grown far too much. Alkali and alkaline earth metals were frequently used in reactions as promoters to speed up the process. Markovic and co-workers reported the direct role of non-covalent interaction of alkali metal cations in hydrogen oxidation reaction (HOR) and oxygen reduction reaction (ORR). [24] Herasymenko et al. made several findings and claimed that the Li<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, and other purportedly "inert" cations had an effect on the kind and concentration of the hydrogen overpotential on mercury electrodes. Shortly after, Tokuoka discovered that putative cation adsorption of (Mg<sup>2+,</sup> Ca<sup>2+,</sup>  $Sr^{2+}$ ,  $Ba^{2+}$ , and  $La^{3+}$ ) [25] significantly altered the electroreduction rates of XO<sub>3</sub><sup>-</sup> type anions (e.g.; BrO<sub>3</sub><sup>-</sup>, NO<sub>3</sub><sup>-</sup>, and IO<sub>3</sub><sup>-</sup>). [26] Li-et-al reported that in various electrode materials, specific alkaline earth metals (e.g., Ba<sup>+2</sup>, Sr<sup>+2</sup>)based cations tend to segregate on the surface, significantly influencing the electrodes catalytic activities. [27] Moreover, Stoffelsma et al. observed

that the effect of Be<sup>2+</sup> on the adsorption phenomena on Pt is comparable to that of Li<sup>+</sup>. [28] Alkali metal cations are thought to indirectly alter the HER kinetics by boosting the Volmer step kinetics. Bandarenka and co-workers reported that HER activity patterns of various Pt electrodes, regardless of their surface shape, are in tune with the hydration energy of the alkali metal cations that is present in the electrolyte, and it shows the following sequence: Li<sup>+</sup> > Na<sup>+</sup> > K<sup>+</sup> > Rb<sup>+</sup> > Cs<sup>+</sup>. [29] Moreover, it is suggested that promoters may modify the hydrogen binding energy (HBE), which further modify the HER activity. In case of Pt catalyst respective to any surface, presence of Li<sup>+</sup> increases HER activity 4 times as compared to Cs<sup>+</sup>. [29] As a result, metal promoters significantly influence the hydrogen's optimal binding energy, which in turn regulates the reaction's thermodynamics. We want to demonstrate in our work how the combination of alkali and alkaline earth metals as promoters affects the metal carbide surfaces for the HER.

#### 1.4. Artificial Intelligence (AI)

Artificial Intelligence (AI) is an important subfield of computer science and engineering dedicated to the creation of intelligent systems capable of performing tasks that traditionally require human intelligence. These tasks encompass a wide range of capabilities, such as visual perception, speech recognition, decision-making and natural language. AI utilizes various techniques including deep learning, machine learning, and natural language processing, and computer vision, among others. The ultimate objective of AI research is to develop machines that possess the ability to learn, reason, and perceive the world in a manner similar to humans. This would enable them to autonomously solve intricate problems and make decisions. The concept of AI was introduced by the renowned British mathematician Alan Mathison Turing in his influential publication "Computing Machinery and Intelligence" in 1950. *[30]* Since then, numerous researchers and scientists have contributed to the advancement of AI. This progress led to the establishment of deep learning (DL) and machine learning (ML) methodologies, which have achieved remarkable accomplishments in the field.



Figure 1.2 Schematic representation of AI, ML and DL domain.

#### **1.4.1. Machine Learning**

Machine learning is a subset of AI that emphases on imparting machines with the ability to learn and enhance their performance through data analysis. This implies that machines can autonomously acquire knowledge and improve based on their experiences, without direct human involvement. Numerous experts and researchers have made significant contributions to the advancement of machine learning. Among them, Arthur Samuel stands out as one of the trailblazers in the field. In 1959, Samuel introduced the notion of ML and gained recognition for creating the initial self-learning program designed for playing checkers.



Figure 1.3 Different types of methods in machine learning

#### 1.4.1.1. Supervised Learning

Supervised machine learning refers to a category of machine learning techniques that utilize labeled datasets for training. In this approach, the machine is exposed to a dataset in which the correct output is already provided, allowing it to learn from these labeled examples. By understanding the relationships between inputs and outputs in the training data, the machine develops a mapping mechanism, which subsequently applies to execute predictions on unknown data. Supervised learning has proven to be valuable in a range of domains, including image classification, natural language processing, and predictive modeling. Supervised machine learning techniques are commonly categorized into two types, regression and classification. In recent studies, regression models in supervised machine learning have been effectively optimized and utilized to forecast the adsorption energies associated with heterogeneous catalysts, which offers a convenient means of screening active and selective catalysts. Furthermore, the adsorption energy is identified as a crucial signifier for the activity of catalysts in previous research studies that have utilized supervised machine learning. [31-35]

#### 1.4.1.2. Features

In ML, the features (also known as input variables or predictors) are the characteristics of the data that are used to make predictions or classifications. The choice of features is critical to the success of a ML model, as the model's performance and accuracy heavily depend on the quality and relevance of the features. Selecting the appropriate features requires understanding of the problem and domain knowledge. The features must be informative and should be able to capture the underlying patterns in the data. Additionally, the features should be independent of each other and not be redundant, as including redundant or correlated features can lead to overfitting and deteriorate model performance. Furthermore, some machine learning algorithms are sensitive to the scale or type of features used. For example, decision trees may not perform well with continuous variables, while neural networks may require normalization of the input data. Thus, it is important to consider the nature of the features and the algorithm being used when selecting and preparing the input data. The choice and quality of features play a crucial role in machine learning, as they directly impact the performance of the model. The selection of informative, independent, and relevant features is essential for achieving optimal model results.

#### **1.4.1.3 ML models:**

There are several types of regression models, including linear regression, kernel ridge regression (KRR), Gradient Boosting Regressor (GBR), eXtreme gradient boosting regression (XGBR) and many more. These regression models can be found and installed from various sources depending on the programming language, tools, and frameworks used. Some common sources include Python packages such as scikit-learn, *[36]* TensorFlow, and Keras, and R packages such as caret, glmnet, and MASS. Cloud-based platforms such as AWS, GCP, and Azure also offer machine learning models as a service or provide tools to build and deploy machine learning models. Standalone software such as SAS and SPSS also offer

regression modeling capabilities. When selecting a regression model, it is essential to consider the nature of the data and the specific problem being addressed, as different models may be more suitable for different situations.

#### • Linear Regression:

Linear regression is a supervised learning technique employed to forecast a continuous output variable by utilizing one or more input variables. It assumes that there exists a linear correlation among the input variables (features), and the output variable. The equation of a linear regression model can be expressed in terms of the following equation:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m \tag{1.1}$$

where y is the predicted output variable,  $b_0$  is the intercept,  $b_1$  to  $b_m$  are the coefficients of the input variables  $x_1$  to  $x_m$ , respectively. The coefficients represent the relation between input and the output variables, and the intercept represents the value of y when all input variables are equal to zero. The objective of linear regression is to determine the optimal coefficients that reduce the disparity among the predicted and observed values of the output variable. This is commonly achieved by minimizing the sum of squared errors between the predicted and observed values through a method known as least squares regression. Linear regression models can be used for various applications, including sales forecasting, weather forecasting, and stock price prediction, among others However, it is crucial to acknowledge that linear regression presupposes a linear correlation between the input and output variables, which may not be appropriate for all kinds of data or situations.

#### Kernel Ridge Regression (KRR):

Kernel ridge regression is a type of regression algorithm that uses the kernel trick to fit a nonlinear model to the data. It is a regularized version of least squares regression that incorporates a penalty term to regulate the complexity of the model and avoid overfitting. The equation for kernel ridge regression can be written as:

$$f(x) = \sum_{i} = \ln \alpha_{i} K(x_{i}, x)$$
(1.2)

where f(x) is the predicted output for a given input x,  $\alpha_i$  are the coefficients for the support vectors, n is the number of support vectors, and  $K(x_i, x)$  is the kernel function that measures the similarity between  $x_i$  and x in the feature space. The kernel function can be chosen based on the features of the data and the problem at hand. Kernel ridge regression seeks to minimize the following objective function:

minimize, 
$$\alpha \|y - X_{\alpha}\|^2 + \lambda \|\alpha\|^2$$
 (1.3)

where  $X_{\alpha}$  is the design matrix, y is the vector of target values,  $\alpha$  is the vector of coefficients. Here, regularization parameter  $\lambda$  controls the tradeoff among fitting the data and keeping the model simple. In summary, kernel ridge regression is a powerful regression algorithm that can handle nonlinear relationships between the input and output variables by modifying the data into a higher-dimensional feature space using a kernel function. The regularization parameter helps to avoid overfitting and increase the model performance in general.

#### • Gradient Boosting Regressor (GBR):

In machine learning, GBR is the abbreviated from of Gradient Boosting Regressor, which is a popular ensemble learning algorithm utilized for regression tasks. Gradient Boosting is a technique where multiple weak prediction models, often referred to as "weak learners" (e.g., decision trees), are combined to create a stronger predictive model. The idea behind gradient boosting is to iteratively train new models that focus on the errors made by the previous models, gradually reducing the overall prediction error. The Gradient Boosting Regressor specifically applies gradient boosting for regression problems. It works by fitting a regression model to the residuals (i.e., the differences among the target values and the predicted values) of the previous model in the ensemble. Each subsequent model is trained to predict the residuals of the previous models, allowing the ensemble to gradually improve its prediction accuracy.

#### eXtreme Gradient Boosting Regression:

XGBoost is classified as an ensemble learning technique, which belongs to a group of methods that merge multiple weak learners (basic models that perform slightly better than random guessing) to create a robust learner. Within the XGBoost framework, decision trees serve as the weak learners, and the algorithm functions by progressively incorporating decision trees into the model. Every new tree aims to rectify the mistakes made by the former trees. The key innovation of XGBoost is that it uses a gradient-based optimization technique to find the best splits in the decision trees. This allows the algorithm to optimize not only the accuracy of the model, but also its complexity, by penalizing overly complex models that are prone to overfitting.

#### 1.4.1.4. Testing and Validation

In machine learning, testing and validation are important steps for evaluating the performance of a model and ensuring that it generalizes well to new data. Testing refers to the process of evaluating the performance of a trained model on an unforeseen dataset. The goal of testing is to estimate the generalized performance of the model on unknown data. Testing is typically done using a separate, held-out dataset that is not used during training. Validation, on the other hand, refers to the process of tuning the hyperparameters of a model for achieving the best possible performance on a validation dataset. The validation dataset is typically a subset of the training dataset that is held out for this purpose. During the validation process, different combinations of hyperparameters are tested, and the one that gives the best performance on the validation dataset is selected. Both testing and validation are important steps in machine learning, as they help to ensure that a model is not overfitting to the training data and can generalize well to new data. It is important to keep in mind that the performance of a model on the training data solely can't indicate its generalizability i.e., the performance of it to new data. Testing and validation provide a more reliable estimate of a model's performance on unseen data. However, it should be noted that these terminologies of test and validation data are sometimes changed purposefully by different authors.

#### 1.4.1.5. Hyperparameter Tunning

Hyperparameter tuning in machine learning involves the selection of optimal hyperparameter values for a given model. Hyperparameters are parameters that are predetermined and not learned from the training data. They are set before the training phase begins, and their values significantly impact the performance and behaviour of the model. They control aspects of the model such as its complexity, regularization, and learning rate, among others. The choice of hyperparameters can be significant for the performance of a ML model, and finding the best set of hyperparameters is often a critical step in building an effective machine learning model. Hyperparameter tuning is typically done using a validation dataset, which is a subset of the training dataset that is used to evaluate the performance of different hyperparameter settings. The process of hyperparameter tuning can be done manually or automatically. There are a number of libraries and tools available for hyperparameter tuning in machine learning, such as scikit-learn, TensorFlow, Keras, and PyTorch. These tools provide a range of options for hyperparameter tuning, from simple grid search to more advanced techniques like Bayesian optimization. Overall, hyperparameter tuning is a critical step in building effective machine learning models, and can have a significant impact on their performance.

### **Chapter-2**

#### **Schrödinger Equation**

The equation proposed by Schrödinger, is a pivotal equation in quantum mechanics. It serves as a mathematical framework to describe the timedependent evolution of a quantum system's wavefunction. This equation is fundamental to our understanding of the behaviour and properties of particles at the quantum level

$$i\hbar\partial\psi/\partial t = H\psi$$
 (2.1)

where  $\psi$ , i, H, t and  $\hbar$  represent the wavefunction of a system, imaginary unit, Hamiltonian operator, time, and reduced Planck constant, respectively. This partial differential equation relates the time derivative of the wavefunction with the Hamiltonian operator acting on the wavefunction. The total energy of the system can be described by a quantum mechanical operator that is Hamiltonian operator, Schrödinger equation depends upon both space and time, and the time independent equation is given by:

$$H\psi = E\psi \tag{2.2}$$

E, represents the total energy of the system. Time-independent Schrödinger equation is used to determine the allowed energy levels of a quantum system. Solving the time-independent Schrödinger equation and Schrödinger equation the is a fundamental problem in quantum mechanics, as it allows us to predict the behaviour of quantum systems and understand their properties. The time-independent Schrödinger equation enables us to find the allowed energy levels of a quantum system, which are important in understanding the behaviour of atoms, molecules, and other quantum systems.

#### **Born-Oppenheimer Approximation:**
The Born-Oppenheimer approximation is a widely used concept in quantum chemistry, which allows for the separation of the nuclear and electronic degrees of freedom in a molecule. In essence, it assumes that speed of electron is faster than speed of nucleus, and thus the electrons can be treated as if they were moving in a static potential generated by the fixed nuclei. Mathematically, this approximation is expressed by the following equation:  $H\psi = E\psi$ , , H represents Hamiltonian operator of the system, which includes both electronic and nuclear contributions. The wave function  $\psi$  is separated into an electronic part  $\Phi$ , which depends on the electronic coordinates only, and a nuclear part X, which depends on the nuclear coordinates only:

$$\Psi = \Phi(r_1, r_2, \dots, r_{N_i}) \times (R_1, R_2, \dots, R_{M_i})$$
(2.3)

where  $r_i$  denotes the electronic coordinates,  $R_i$  denotes the nuclear coordinates, and N and M represent the number of electrons and nuclei, respectively.

By using this separation of variables, the Hamiltonian can be rewritten as a sum of electronic and nuclear Hamiltonians:  $H = H_e + H_n$ . Where  $H_e$  is the electronic Hamiltonian, which depends on the electronic coordinates only, and  $H_n$  is the nuclear Hamiltonian, depends solely on the nuclear coordinates. This approximation easily calculate total wave function and energy of a system even for larger molecules

#### Hartree-Fock (HF) Approximation

Hartree-Fock (HF) approximation is a technique used to solve the electronic structure problem of a molecule or atom in quantum chemistry. It assumes that the electrons move independently in an effective potential created by the average electron density of various other electrons. The HF wave function can be written in terms of single Slater determinant, which is a function of the molecular orbitals (MOs) that are created by linear combinations of atomic orbitals (LCAOs). The solution of Schrödinger

equation with the effective potential gives MOs and are chosen to minimize the system energy. The HF energy is the addition of kinetic energy of the electrons and the electron-electron Coulomb repulsion energy. This is known as the HF energy expression:

$$E(HF) = \sum i \sum j \left\langle \Psi_i \right| \frac{-1}{2\nabla 2} \left| \Psi_j \right\rangle + \sum i \sum j \left\langle \Psi_i \Psi_j \right| \frac{1}{|r_1 - r_2|} \left| \Psi_j \Psi_j \right\rangle$$
(2.4)

where  $\Psi_i$  and  $\Psi_j$  are the MOs,  $r_1$  and  $r_2$  are the positions of two electrons, and the brackets denote integration over all space. The HF method provides a good initialization for more accurate methods that include electron correlation effects beyond the HF approximation. However, its simplicity and efficiency make it widely popular in calculating the electronic structure of small and medium-sized molecules.

## **Density Functional Theory (DFT)**

Density Functional Theory (DFT) used in quantum chemistry to solve the solid or molecular electronic structure problem. According to it, functional of the electron density can express the total energy of the system rather than the more complicated wave function, which simplifies the calculations significantly. The Kohn-Sham equation plays central role in DFT, which defines a set of auxiliary non-interacting electrons that exhibit electron density that is at par with that of true interacting electrons. The Kohn-Sham equation is mentioned below:

$$\left[-\frac{1}{2\nabla 2} + v(r) + v_H(r) + v_{xc}(r)\right]\phi_i(r) = \varepsilon_i\phi_i(r)$$
(2.5)

Where  $\phi_i(r)$ , v(r),  $v_H(r)$ , and  $v_{xc}(r)$  are the Kohn-Sham orbitals, external potential, Hartree potential because of the electron-electron Coulomb repulsion, and exchange-correlation potential that accounts for the electronelectron correlation effects, respectively. Moreover,  $\varepsilon_i$  are the Kohn-Sham eigenvalues, which correspond to the energy levels of the auxiliary electrons. The sum of all occupied Kohn-Sham orbitals gives the electron density which can be calculated with the following equation (2.6)

$$\mathbf{p}(r) = \sum_{i} |\phi_{i}(r)|^{2}$$
(2.6)

 $v_{xc}(r)$  is typically approximated using an electron density dependent functional. The most widely used functional is the local density approximation (LDA), which presumes that the exchange-correlation potential is a function of the electron density at every point in space. DFT is the most widely used methods in quantum chemistry and solid-state physics due to its efficiency and accuracy in computation of the electronic structure of large systems. However, the selection of the exchangecorrelation functional can bear a substantial impact on results, and more advanced functionals are being developed to make DFT calculations more efficient.

#### **The Hohenberg-Kohn Theorems**

The Hohenberg-Kohn theorems are fundamental theorems of DFT that establish the mathematical foundation of the theory. There are two theorems: The Hohenberg-Kohn's first theorem asserts that the ground-state electron density has an effect on the external potential of a system uniquely. Mathematically, this theorem can be expressed as: For any two external potentials,  $V_1(r)$  and  $V_2(r)$ , that produce identical ground-state electron density  $\rho(r)$ , the two potentials must be identical up to an additive constant:

$$V_1(r) = V_2(r) + C$$
, where C is a constant. (2.7)

The Hohenberg-Kohn's second theorem states that the ground-state energy of a system is a unique functional of the electron density. Mathematical expression of the theorem is as follows: The ground-state energy of a system,  $E[\rho]$ , is a unique functional of the electron density,  $\rho(r)$ , and is determined by energy functional's minimum value,  $E[\rho]$ , over all possible electron densities that have the same normalization and external potential:

$$E[\rho] = \min \{F[\rho]\}$$
(2.8)

where  $F[\rho]$  is the energy functional that depends on the external potential and the electron density, and is calculated as:

$$\mathbf{F}[\rho] = \sum i \,\varepsilon_i \,n_i - \int p(r) \left[ v(r) + v_H(r) + v_{xc}(r) \right] \,\mathrm{dr} \tag{2.9}$$

Where  $\varepsilon_i$ , n<sub>i</sub>,  $v_H(r)$ ,  $v_{xc}(r)$ , are the measures of Kohn-Sham eigenvalues, occupation numbers, Hartree potential due to the electron-electron Coulomb repulsion, and the exchange-correlation potential that accounts for the electron-electron correlation effects respectively. v(r) is the measure of the external potential. These theorems provide the basis for the practical implementation of DFT to mathematically compute the electronic structure of molecules and materials.

#### **Kohn-Sham Formulation**

The Kohn-Sham formulation is a widely used method in DFT for approximating the ground-state electronic structure of a many-electron system. The basic idea behind the Kohn-Sham method is to potentially substitute the interacting system with a non-interacting system with similar electron density. This non-interacting system is then solved using standard techniques from quantum mechanics. These equations are a set of selfconsistent equations that are used to compute the system's electronic properties. The equations are derived by introducing a set of fictitious noninteracting particles that occupy the same electronic density of interacting system. These particles are referred to as Kohn-Sham orbitals, and they satisfy a set of equations that are similar to the Schrödinger equation. The Kohn-Sham equations are given by:

$$\left(-\frac{1}{2}\right)\nabla^{2}\Psi_{i}(r) + V_{eff}(r)\Psi_{i}(r) = \mathcal{E}_{i}\Psi_{i}(r)$$
(2.10)

where  $\Psi_i(r)$  is the value of i-th Kohn-Sham orbital,  $\varepsilon_i$  denotes energy of the i-th Kohn-Sham orbital, and  $V_{eff}(r)$  denotes effective potential that

includes the exchange-correlation potential and the external potential. The effective potential  $V_{eff}(r)$  can be computed with the help of following expression:

$$V_{eff}(r) = V_{ext}(r) + \int \frac{\rho(r')}{|r-r'|} dr' + V_{xcp}$$
(2.11)

Where  $V_{ext}(r)$  is the external potential,  $\rho(r)$  is an electron density, and  $V_{xcp}$  is the exchange-correlation potential that accounts for the exchange and correlation interactions between electrons. These equations are solved self-consistently, meaning that the electron density is assessed with the assistance of the Kohn-Sham orbitals, and then the new electron density is used to recalculate the effective potential. This process is continued to achieve the convergence. The following approximations are widely employed in simulating both molecular and solid-state problems. [37]

#### **Exchange-correlation Functional**

Exchange-correlation functional in DFT is a term that accounts for the exchange and correlation interactions between electrons in a many-electron system. It is typically denoted as  $E_{xc}$  ( $\rho$ ), where  $\rho$  represents the electron density, and it can be parted into two functional components: the exchange and the correlation. The most commonly used XC functional is LDA.

## Local Density Approximation (LDA).

In DFT, the method of Local density approximation (LDA) assists in approximating the XC energy functional by assuming that it solely depends on the electron density at every point in the space. The LDA is the simplest form of XC functional and is widely used in DFT calculations. It supposes that the XC energy density and  $\varepsilon_{xc}$  is a function of the electron density  $\rho[(\mathbf{r})]$  at each point in space:

$$\varepsilon_{xc} \left[ \rho(\mathbf{r}) \right] = f[\rho(\mathbf{r})] \tag{2.12}$$

Where the electron density is the sole determinant of function f. The total exchange correlation energy  $E_{xc}$  is calculated by integrating the XC energy density over all space:

$$E_{xc} = \int \varepsilon_{xc}[p(r)]d^3r \qquad (2.13)$$

by using the LDA, the XC energy can be computed with the help of following equation:

$$E_{xc} = \int p(r) \varepsilon_{xc}[p(r)] d^3 r = \int p(r) f[p(r)] d^3 r$$
 (2.14)

Where, p(r) represents electron density at every point in space. The LDA can be described as an approximation that neglects the non-local effects of the XC functional, which can lead to errors in the calculated electronic properties. However, it is computationally efficient and can provide reasonable results for many systems.

## **Generalized Gradient Approximation (GGA)**

The generalized gradient approximation (GGA) is a procedure for approximating the exchange-correlation (XC) energy functional in DFT by considering the electron density and its gradient. The GGA is more accurate than the LDA, but it is more computationally expensive. The Perdew-Burke-Ernzerhof (PBE), Perdew and Wang (PW91), PBEsol and revised PBE (RPBE) functionals are mostly used GGA functionals in DFT calculations.

## **Basis Sets**

Basis sets can be described as a set of mathematical functions which assist to approximate the wave function of a molecule or system. Gaussian-type orbitals (GTOs) and Slater-type orbitals (STOs) are two commonly used orbitals. STOs are more complex but more flexible and can better describe the behaviour of electrons close to the nucleus, while GTOs are simpler but less flexible and can only approximate the behaviour of electrons near to the nucleus. Other types of basis sets include polarized, diffuse, and mixed basis sets, which combine STOs and GTOs to balance the advantages and disadvantages of each type. Basis sets are essential for accurately calculating molecular properties such as energy, structure, and spectroscopic properties in quantum chemistry.

#### **Projector Augmented Wave (PAW) Method**

The Projector Augmented Wave(PAW) is a technique used in computational studies and quantum chemistry for accurately modeling the electronic structure of various complex systems, like solids and surfaces. The PAW method merges the accuracy of all-electron methods with the efficiency of pseudopotential methods. In this method, the true all-electron wave function is replaced with a pseudo wave function that is expanded in a set of auxiliary basis functions. The pseudo wave function is constructed such that it has the similar core electron density to the true all-electron wave function, while the valence electron density is modified to account for the core-valence interaction. This allows for the use of a smaller basis set for the valence electrons, that minimize the computing cost of the calculations.

The pseudo wave function is defined as:

$$\Psi_{p}(r) = \Psi_{v}(r) + \Sigma \Psi_{c}(r) \Phi r \qquad (2.15)$$

where  $\Psi_{v}(r)$  is the true valence electron wave function,  $\Psi_{c}(r)$  is the core electron wave function, and  $\Phi r$  is the smooth step function that smoothly transitions from 0 to 1 as r approaches the atomic radius. The projection operator, P(r), is equated as:

$$P(r) = \Psi_{\nu}(r)\Psi_{\mu}(r)/\Psi_{\nu}(r)\Psi_{\nu}(r)$$
(2.16)

The PAW method uses the projection operator to project the density and potential from the all-electron space to the pseudo space and back. The overall energy of the system is calculated as the addition of total kinetic energy, potential energy, and electron-electron repulsion energy. The PAW method has become a popular approach in both materials science and quantum chemistry due to its ability to accurately model complex systems while maintaining computational efficiency

#### **GPAW** and **ASE**

GPAW (Grid Based Projector Augmented Wave) is a software program that used PAW (Projector Augmented Wave) method for first principle calculations. It can be used in combination with the Atomic Simulation Environment (ASE) and is open source. Users interact with GPAW through command scripts, which are primarily written in the programming language Python. [38,39]

#### **Computational Details.**

GPAW algorithm in the projector augmented wave (PAW) method is utilized for all the DFT calculations. The exchange-correlation potential, implemented in the atomistic simulation environment (ASE), was chosen to use the Perdew-Burke-Ernzerhof generalized gradient approximation (GGA–PBE) [40,41] functional. In order to calculate the adsorption energy and conduct additional studies, periodic surface slabs  $(4 \times 4 \times 4)$  having 4 layers for (100) and (111) surfaces were taken into account. For the planewave expansion of the electronic wave function, a cut-off of 400 eV has been taken. Iterative diagonalization of the Kohn-sham Hamiltonian calculations has used the Davidson (DAV) eigen solver method with the Monk horst-pack k-point sampling of the Brillouin zone of (3,3,1). To eliminate any periodic image interactions, a ~15Å vacuum was added to the catalyst surface in the z-direction. During the geometry optimization process, the upper and the lower two layers were allowed to relax and fix, respectively and a convergence criterion of 0.09 eV/Å was used for force minimization.



**Figure 2.1** The  $4 \times 4 \times 4$  periodic slab of Hafnium carbide with (100) surface. a) Top view and, b) side view. The crossed spheres of the lower two layer signifies the fixed atoms, and uncrossed spheres represents the relaxed atoms for the adsorption energy calculations. Cyan colour represents Hf atoms and grey colour represents carbon atoms.



**Figure 2.2** The 4×4×4 periodic slab of Hafnium carbide with (111) surface in which metal is present on the top layer and followed by carbon layer. a) Top view and, b) side view. The crossed spheres of the lower two layer signifies the fixed atoms, and uncrossed spheres represents the

relaxed atoms for the adsorption energy calculations. Cyan colour represents Hf atoms and grey colour represents carbon atoms.



**Figure 2.3** The  $4\times4\times4$  periodic slab of Hafnium carbide with (111) surface in which carbon is present on the top layer and followed by metal layer. a) Top view b) side view. The crossed spheres of the lower two layer signifies the fixed atoms, and uncrossed spheres represents the relaxed atoms for the adsorption energy calculations. Cyan colour represents Hf atoms and grey colour represents carbon atoms.

The adsorption energy of H ( $\Delta E_H$ ) has been computed using the following equation:

$$(\Delta E_H) = E_{sp*} - (E_{sp} + E_{H*})$$

 $E_{sp*}$  represents the corresponding DFT energy of the adsorbed surface when a promoter is present, $E_{sp}$  is merely the surface and promoter's DFT energy and energy of adsorbed H is denoted by  $E_{H*}$ .

#### **Result and Discussions**

Firstly, we created 64 atom slabs consisting of different transition metal carbides, with each slab containing 16 atoms arranged in four layers (4x4x4). The top two layers of these 64 atom slabs were allowed to relax during the reaction, while the other two layers were kept fixed. To accomplish this, we generated a set of 81 structures using computed DFT values as training data. Subsequently, we determined the importance of each feature in relation to the output data (adsorption energy) to gain insight into their relationship. We, then, optimized several supervised regression models through hyperparameter tuning using the training data and evaluated their performance with the calculation of test data in order to identify the best model for prediction. Lastly, we screened for active catalysts by considering and comparing the adsorption energies of total 1485 structures, with the adsorption energies of the pure platinum metal carbide slab model and similar systems.

## **3.1.**Catalyst Modelling

To analyze the adsorption of H atom on different transition metal carbides, we focused on two types of surfaces, such as (100) and (111). These surfaces exhibit different arrangements of carbon and transition metals. In the (100) surface, both the metal and carbon atoms reside on the same layer, leading to a single type of available hollow site for adsorption, known as four-fold hollow (ffh). On the other hand, the (111) surface consists of alternating layers of metal and carbon, resulting in two types of top configurations: metal on top and carbon on top. Consequently, two types of hollow sites are present: hollow fcc and hollow hcp. Firstly we considered different hollow sites of all the chosen catalytic surfaces for H adsorption. To enhance the efficiency of the HER, we introduced alkali and alkaline earth metals as promoters in nearby hollow sites of H atom. These promoters interact electrostatically with the TMCs. Considering different combinations of alkali, alkaline earth, and pure promoters, we obtained a

total of 1485 possible catalyst-promoter combinations. However, even with the fastest ab initio tools available, calculating DFT values for such a large number of combinations is practically infeasible. This limitation is the principal reason for employing machine learning (ML) techniques.



**Figure 3.1** (a) and (b) are the Top and side view of (100) TMC surface, where cyan colour represents Hf and grey colour represent carbon. (c) represents, adsorption of H atom in a four-fold hollow (ffh) site present in (100) surface and (d) represents, adsorption of promoters near the adsorbed hydrogen in the hollow site of metal carbide (100) surface. Here, cross sign indicates fixed atoms.



**Figure 3.2** (a) and (b) are the Top and side view of (111) TMC surface, when metal is present on the top and cyan colour represents Hf and grey colour represent carbon. (c) and (d) represents, adsorption of H atom in fcc and hcp hollow site present in (111) surface, whereas (d) and (e) represents, adsorption of promoters near the adsorbed hydrogen in the fcc and hcp hollow sites of metal carbide (111) surface. Here, cross sign indicates fixed atoms.



**Figure 3.3** (a) and (b) are the Top and side view of (111) TMC surface, when carbon is present on the top and cyan colour represents Hf and grey colour represent carbon. (c) and (d) represents, adsorption of H atom in fcc and hcp hollow site present in (111) surface, whereas (d) and (e) represents, adsorption of promoters near the adsorbed hydrogen in the fcc and hcp hollow sites of metal carbide (111) surface. Here, cross sign indicates fixed atoms.



## **3.2.Feature Selection**

**Figure 3.4** Graphical representation of different sets of input features. Each set consists of different elemental features of carbon, metal and promoters .

Feature importance reveals the weighting assigned to each descriptor by the model. The selection of input features plays a crucial role in ML-based studies. By constructing and training a machine using these input features and their corresponding output data, we can predict outcomes for unknown data, significantly reducing computational costs. In this work, our objective is to investigate an approach for predicting the adsorption energy of unidentified materials in the context of the HER. Additionally, based on the metals under consideration, we aim to identify active and selective catalysts. By examining the characteristics (chemical properties) of 3d, 4d, and 5d

TMCs and promoters, we gain computational insights into the relationship between these parameters and the HER. When employing alkali and alkaline earth metals as promoters for TMCs, we can better understand how compositional parameters influence the reaction. These insights provide valuable chemical knowledge regarding the most important traits for predicting synthesizability data. Since, TMCs consist of both transition metals and carbon, the features of both components are equally significant. Therefore, certain compositional and elemental characteristics hold particular importance. We consider various explanatory variables (chemical properties) such as atomic mass, atomic radius, Pauling electronegativity, electron affinity, valence electrons, Mendeleev number, first ionization energy, dipole polarizability, and Martynov-Batsanov electronegativity for the transition metal, carbon, and promoter. We also consider the coordination numbers of both (100) and (111) metal surfaces. In the case of the metal-on-top (111) surfaces, we calculate the feature difference between the metal and promoter for all explanatory variables, and similarly for the carbon-on-top surfaces. When both carbon and metal are present on the surfaces, as in the case of (100) surfaces, we calculate the difference between average of the promoters and the metal of TMCs as an explanatory variable. These characteristics enable us to differentiate between various constituent transition metal carbides with different promoter combinations and understand their electronic behavior. Among these explanatory variables, we must identify those that contribute most significantly to the output, specifically the adsorption energy. To accomplish this, we employ machine learning techniques, primarily supervised machine learning, to train a model using these explanatory variables.

**Table 3.1** List of all the selected input features

Sl. No.	Considered Features	SI. No.	Considered Features
0.	Coordination Number	17	Martynov- Batsanov Electronegativity(P1)
1.	Lattice parameter a	18	Mass(P1)
2.	Lattice parameter b	19	Atomic Radii(P1)
3.	Lattice parameter c	20	Valence Electron(P1)
4.	Lattice parameter A	21	First Ionization Energy(P1)
5.	Lattice parameter B	22	Dipole Polarizability(P1)
6.	Lattice parameter C	23	Mendeleev Number(P1)
7.	Pauling Electronegativity(M/C)	24	Effective Nuclear Charge(P1)
8.	Martynov- Batsanov Electronegativity(M/C)	25	Pauling Electronegativity(P2)
9.	Mass(M/C)	26	Martynov- Batsanov Electronegativity(P2)
10.	Atomic Radii(M/C)	27.	Mass(P2)
11.	Valence Electron(M/C)	28.	Atomic Radii(P2)
12.	First Ionization Energy(M/C)	29.	Valence Electron(P2)
13.	Dipole Polarizability(M/C)	30.	First Ionization Energy(P2)
14.	Mendeleev Number(M/C)	31.	Dipole Polarizability(P2)
15.	Effective Nuclear Charge(M/C)	32.	Mendeleev Number(P2)
16	Pauling Electronegativity(P1)	33	Effective Nuclear Charge(P2)

SI. No.	Considered Features	SI. No.	Considered Features
34.	Pauling Electronegativity(SurfaceM/C- P1)	52.	Martynov- Batsanov Electronegativity(M/C)
35.	Martynov- Batsanov Electronegativity(SurfaceM/C- P1)	53.	Mass(M/C)
36.	Mass(SurfaceM/C-P1)	54.	Atomic Radii(M/C)
37.	Atomic Radii(SurfaceM/C- P1)	55.	Valence Electron(M/C)
38.	Valence Electron(SurfaceM/C- P1)	56.	First Ionization Energy(M/C)
39.	First Ionization Energy(SurfaceM/C- P1)	57.	Dipole Polarizability(M/C)
40.	Dipole Polarizability(SurfaceM/C- P1)	58.	Mendeleev Number(M/C)
41.	Mendeleev Number(SurfaceM/C-P1)	59.	Effective Nuclear Charge(M/C)
42.	Effective Nuclear Charge(SurfaceM/C- P1)	60.	Pauling Electronegativity(SurfaceM/C- Promoter Average{P.A})
43.	Pauling Electronegativity(SurfaceM/C- P2)	61.	Martynov- Batsanov Electronegativity(SurfaceM/C- P.A)
44.	Martynov- Batsanov Electronegativity(SurfaceM/C- P2)	62.	Mass(SurfaceM/C-P.A)
45.	Mass(SurfaceM/C-P2)	63.	Atomic Radii(SurfaceM/C- P.A)
46.	Atomic Radii(SurfaceM/C- P2)	64.	Valence Electron(SurfaceM/C- P.A)
47.	Valence Electron(SurfaceM/C- P2)	65.	First Ionization Energy(SurfaceM/C- P.A)
48.	First Ionization Energy(SurfaceM/C- P2)	67.	Dipole Polarizability(SurfaceM/C- P.A)
49.	Dipole Polarizability(SurfaceM/C- P2)	68.	Mendeleev Number(SurfaceM/C-P.A)
50.	Mendeleev Number(SurfaceM/C-P2)	69.	Effective Nuclear Charge(SurfaceM/C- P.A)
51.	Effective Nuclear Charge(SurfaceM/C- P2)		
52.	Pauling Electronegativity (PA)		



**Figure 3.5** Representation of the feature importance of datasets containg (111) and (100) surfaces having hcp and ffh hollow sites by considering 28 input Features of carbon, metal and promoters.



**Figure 3.6** Representation of the feature importance of datasets containg (111) and (100) surfaces having fcc and ffh hollow sites by considering 70 input Features of carbon, metal and promoters

In this study, we have considered two types of surfaces such as (100) and (111) surfaces. In the case of (100) surfaces, both metal and carbon exist on the top layer where adsorption takes place. However, in the case of (111) surfaces, metal and carbon alternate between layers. Initially, catalytic surfaces have been modelled for hydrogen adsorption at the hollow site, as previously discussed. Then, the reaction's efficiency enhancement by introducing promoters from alkali and alkaline earth metals in nearby hollow sites in both (100) and (111) surfaces has been assessed. For this, ML algorithms are employed, and various input features are considered. For instance, in the case of (100) surface, when both carbon and metal are present on top, we have considered the average of their properties to define the surface properties. In the case of (111) surface, we considered carbon and metal-based features separately when there are corresponding carbonon-top or metal-on-top structures, respectively. Additionally, we evaluate the promoters' features individually and combined by averaging them to determine whether the individual promoter property has a greater influence. Upon analyzing the feature importance plot from ML, we have found that the input feature such as dipole polarizability of (metal – promoter average), dipole polarizability of individual metal and carbon, mass of individual carbon and metal and atomic radii of metal and carbon dominate in both of the considered datasets. The dominance of dipole polarizability as a feature in TMCs for the HER is linked to their electronic structure. When a promoter is adsorbed in the vicinity of the transition metal carbide, the dipole moment of the promoter induces an electrostatic interaction which can influence the electronic structure and reactivity of the carbide surface, thus affecting the HER activity and also the presence of a promoter with a significant dipole moment can induce charge. This charge redistribution can lead to a modification of the electronic density and local bonding environment at the catalytic sites, which in turn affects the HER activity that's why dipole polarizability influenced the most in case of individual elemental features of carbon and metal or in a combined way with the

promoters as in our both the datasets. When it comes to mass, the adsorption of hydrogen species on TMC surfaces involves the electrostatic attraction between the hydrogen atoms and the active sites. The mass-promoting elements, such as transition metals with higher atomic masses, can provide additional binding sites and increase the overall binding energy of the adsorbed species. This higher binding energy enhances the stability of the intermediates involved in the HER and lowers the energy barriers for the reaction. As we have taken 9 different types of transition metal carbides having different masses so carbides with larger masses influenced more. In case of promoters the introduction of alkali or alkaline earth metal atoms into the TMCs surface can induce charge subsequently reorients electronic properties, enhancing the adsorption and reactivity of hydrogen species. In TMCs, the metal atom forms a strong bond with the carbon atoms. The bond length and bond strength are influenced by the atomic radii of the participating atoms. As the atomic radii of the transition metal atoms increases, the metal-carbon bond length generally increases. This longer bond length weakens the metal-carbon bond, making it easier to break and facilitate the release of hydrogen gas during the HER. Overall, all these features have significant physicochemical influence with the output adsorption energy and thus can be treated as the important parameters for the prediction of H adsorption energy in HER.

## **3.3.ML Model Selection**

We have considered three supervised machine learning regressors for the prediction task. We explored the utilization of kernel-based and gradient boosting framework algorithms, specifically Kernel Ridge Regression (KRR) and Gradient Boosting Regression (GBR), which are implemented in the Scikit-learn package. Additionally, we employed a gradient boosting framework called eXtreme gradient boosting regression (XGBR). All of these algorithms are readily available in their respective open-source libraries. In our study, we applied these algorithms to two datasets. The first

dataset containg (111) and (100) surfaces, which includes four-fold hollow site and hcp site for the adsorption of hydrogen. The second dataset involves (111) and (100) surfaces, which also consists a four-fold hollow site but an fcc hollow site for adsorption.

# **3.4. Model Optimization**

These algorithms which come with default parameters, can be fine-tuned to better suit the user's needs. In our study, we adjusted the hyperparameters of these algorithms using two commonly employed techniques: Randomized Search CV and Grid Search CV, both implemented in the Scikit-learn package. By applying these techniques, we were able to find the optimal values for the hyperparameters, resulting in highly optimized algorithms.

To assess the effectiveness of each model, we conducted training and testing using separate test and train datasets. Subsequently, we assessed the Mean Absolute Error (MAE) and also Root Mean Square Error (RMSE) for each model, which is a common evaluation metric in machine learning. As is typical in ML, for a given dataset, one of the optimized algorithms usually outperforms the others, exhibiting the lowest RMSE and MAE values providing the best predictions. Therefore, based on the RMSE and MAE values obtained for all the considered optimized regressors, we selected the best individual models for each dataset. These models were then used for the final prediction, as they demonstrated superior performance compared to the others.

## **3.5.Prediction Evaluation**

Among the various algorithms considered, we observed that the optimized Kernel Ridge Regression (KRR) achieved the lowest test and train RMSE and MAE values for datasets containing hcp and ffh hollow sites, as well as for datasets containing fcc and ffh hollow sites. the optimized eXtreme gradient boosting regression (XGBR) showed better performance in terms of test RMSE and MAE. However, it's important to note that the XGBR model exhibited very low training RMSE and MAE, indicating potential overfitting issues. On the other hand, the KRR model displayed a scientifically acceptable training RMSE and MAE.

Considering these factors, we decided to utilize the optimized KRR model for both datasets, as it demonstrated superior performance in terms of test RMSE and MAE and avoided overfitting concerns.

The range of DFT-calculated adsorption energies in both datasets was relatively narrow. However, as mentioned earlier, we expanded our analysis to include a well-sampled dataset consisting of 55 combinations, which represented a subset of the complete 1485 combinations. We assessed the performance of our models and found them to be satisfactory and acceptable, particularly for the datasets containg fcc and ffh hollow sites. Among the models, the Kernel Ridge Regression (KRR) model performed relatively better than XGBR and GBR in this dataset.



**Figure 3.7** Plot of predicted adsorption energies ( $\Delta$ Epred) versus DFT calculated adsorption energies ( $\Delta$ Ecalc) using optimized KRR model for the datasets containing (111) surfaces with fcc hollow sites and 100 surfaces with ffh



**Figure 3.8** Plot of predicted adsorption energies ( $\Delta$ Epred) versus DFT calculated adsorption energies ( $\Delta$ Ecalc) using optimized KRR model for the

datasets containing (111) surfaces with hcp hollow sites and 100 surfaces with ffh

**Table 3.2** Optimization of hyperparameters with their corresponding RMSE and MAE values for datasets containing (111) and (100) surfaces having fcc and ffh sites with the three considered regression models KRR, GBR and XGBR.

ML	Optimized	RMSE(eV)	MAE(eV)
models	Hyperparameters		
KRR	kernel=laplacian, gamma=0.00001,	0.34	0.27
	degree=3, coef0=0, alpha=0.01		
GBR	n_estimaters=1100,	0.24	0.14
	min_samples_split = 5, max_depth = 5		
	min_samples_leaf = 2, learnig_rate =		
	0.2		
XGBR	n estimators =1100.	0.18	0.19
nobri	min_child_weight = 2, max_depth = 5,		
	learning_rate = 0.15, booster = gotree,		
	base_score = 1		

**Table 3.3** Optimization of hyperparameters with their correspondingRMSE and MAE values for datasets containing (111) and (100) surfaces

having hcp and ffh sites with the three regression models KRR, GBR and XGBR.

ML	Optimized Hyperparameters	RMSE(eV)	MAE(eV)
models			
KRR	kernel = laplacian, gamma = 0.001,	0.35	0.29
	degree = 2, $coef0 = 4$ , $alpha = 0.1$		
GBR	n_estimators=1100, min_samples_split	0.50	0.39
	= 5, min_samples_leaf = 2, max_depth		
	= 5, learning_rate = 0.2		
XGBR	n_estimators =1100, min_child_weight	0.39	0.28
	= 1, max_depth = 10, learning_rate =		
	$0.15$ , booster = gbtree, base_score = 1		

For the datasets containg fcc and ffh hollow sites, the test RMSE and MAE was lower for GBR and XGBR compared to KRR. However, we observed that the training RMSE and MAE for GBR and XGBR was extremely low, which is not considered acceptable due to the potential issue of overfitting. Consequently, we determined that the KRR model was the best fit

optimized model for both the for datasets containing hcp and ffh hollow sites, as well as for datasets containing fcc and ffh hollow sites.

## **3.6 Stability**

To check whether our model work with unseen data we did cross validation for KRR. Cross-validation is a technique used in machine learning (ML) to assess the performance and generalization ability of a model on unseen data. It involves partitioning the available dataset into various subsets, or "folds," and iteratively evaluating and training the model on different combinations of these folds. For this we have divided our both for datasets containing hcp and ffh hollow sites, as well as for datasets containing fcc and ffh hollow sites into 5 mutually exclusive folds labelled as fold1, fold2, fold3, fold4, fold5. Firstly we use fold2, fold3, fold4, fold5 as our training datasets and then check the model's performance on fold1 (the validation set) and record the evaluation metrics. Then in next we take fold1, fold3, fold4, fold as training datasets and again check the model's performance on fold2 this time. we did this thing for 5 times for each fold and record its cross score. Actually this helps in identifying overfitting, which generally happen when a model performs well on the training data but poorly on newly data. By testing the model on various parts of the data, cross-validation helps us determine if the model is too focused on the specific details of the training set.



**Figure 3.9** Representation of MAE scores for datasets (111) and (100) surfaces having hcp and ffh hollow sites, at 5 fold cross validation



**Figure 3.10** Representation of MAE scores for datasets (111) and (100) surfaces having fcc and ffh hollow sites, at 5 fold cross validation

# 3.7. Catalysts screening

Our work involves using TMCs as catalysts for the HER instead of the currently available Pt benchmark catalyst. To screen catalysts, we set the screening criteria using the equation

$$[\Delta E_H^{Pt} - 1] < \Delta E_H^x < [\Delta E_H^{Pt} + 1]$$

over 1485 adsorption energy values. The screening criteria range was set at  $(\Delta E_H^{Pt} \pm 1)$ , and we will further study the catalysts that fall within this range. However, calculating the DFT values for transition metal carbides is timeconsuming, and currently, we only have 60 out of the previously decided 81 datasets. Therefore, we need to calculate the remaining 21 datasets in the future and compare them with Pt carbide for the final result. We will use the same DFT tool (GPAW) and computational conditions on a pure Pt(111) and Pt(100) slab of 64 atoms (4×4×4) with the considered lattice constants.

Surface	On top atoms	Promoters	Adsorption Energy
HfC (100)	Hf, C	(Li, Cs)	-0.75
HfC (100)	Hf, C	(Rb, Mg)	-0.45
HfC (100)	Hf, C	(Rb, Ba)	-0.57
CrC (100)	Cr, C	(Ca, Cs)	-0.27
CrC (100)	Cr, C	(Li, Rb)	-0.23
NbC (100)	Nb, C	(Na, Rb)	-0.52
NbC (100)	Nb, C	(Sr, Ba)	-0.60
NbC (100)	Nb, C	(Na, Ca)	-0.78
TiC (100)	Ti, C	(Be, Be)	-1.00
TiC (100)	Ti, C	(Li, Be)	-0.40
TiC (100)	Ti, C	(K, Mg)	-0.33
TaC (100)	Ta, C	(Na, Sr)	-0.68
TaC (100)	Ta, C	(Cs, Mg)	-0.53
TaC (100)	Ta, C	(K, K)	-0.72
ZrC (100)	Zr, C	(K, Ba)	-0.71

**Table 3.4** Adsorption energies of different combination of TMCs andpromoter for 100 surfaces containing ffh site.

Surface	On top atoms	Promoters	Adsorption
			Energy (eV)
ZrC (100)	Zr, C	(Sr, Sr)	-0.59
ZrC (100)	Zr, C	(Li, K)	-0.74
ZrC (100)	Zr, C	(Li, Ba)	-1.06
VC (100)	V, C	(Rb, Cs)	-0.08
VC (100)	V, C	(Mg, Ba)	-0.19
VC (100)	V, C	(Be, Ca)	-1.05
VC (100)	V, C	(Na, Cs)	-0.15
WC (100)	W, C	(Li, Ba)	-0.93
WC (100)	W, C	(Be, Ba)	-1.00

**Table 3.5** Adsorption energies of different combination of transition metal carbides and promoter for (111) surfaces in which metal is on top containing fcc hollow site.

Surface	On top atoms	Promoters	Adsorption Energy (eV)
CrC (111)	Cr	(Ca, Sr)	-0.64
CrC (111)	Cr	(K, Mg)	-0.73
CrC (111)	Cr	(Rb, Mg)	-0.74
CrC (111)	Cr	(Li, Sr)	-0.66
WC (111)	W, C	(Ca, Cs)	-1.19
WC (111)	W, C	(Rb, Rb)	-1.06
WC (111)	W, C	(K, Cs)	-1.07
HfC (111)	Hf	(Ca, Be)	-1.13
HfC (111)	Hf	(Mg, Ba)	-1.42
VC (111)	V	(K, Ca)	-1.01
VC (111)	V	(Mg, Mg)	-1.02
VC (111)	V	(Rb, Be)	-0.75
VC (111)	V	(K, K)	-1.01
MoC (111)	Мо	(Na, K)	-1.74
NbC (111)	Nb	(Ca, Ba)	-1.09
TaC (111)	Та	(Na, Ba)	-1.12
TaC (111)	Та	(Mg, Ca)	-1.17

Surface	On top atoms	Promoters	Adsorption Energy (eV)
TiC (111)	Ti	(Cs, Be)	-0.91
ZrC (111)	Zr	(Cs, Sr)	-1.04

**Table 3.6** Adsorption energies of different combination of transition metal carbides and promoter for (111) surfaces in which metal is on top containing hcp hollow site.

Surface	On top atoms	Promoters	Adsorption Energy (eV)
CrC (111)	Cr	(Ca, Sr)	-0.35
CrC (111)	Cr	(K, Mg)	-0.44
CrC (111)	Cr	(Rb, Mg)	-0.46
CrC (111)	Cr	(Li, Sr)	-0.39
WC (111)	W, C	(Ca, Cs)	-1.19
WC (111)	W, C	(Rb, Rb)	-1.06
WC (111)	W, C	(K, Cs)	-1.07
HfC (111)	Hf	(Ca, Be)	-0.59
HfC (111)	Hf	(Mg, Ba)	-0.68
VC (111)	V	(K, Ca)	-0.79
VC (111)	V	(Mg, Mg)	-0.52
VC (111)	V	(Rb, Be)	-0.52
VC (111)	V	(K, K)	-0.78
MoC (111)	Мо	(Na, K)	-1.24
NbC (111)	Nb	(Ca, Ba)	-0.69
TaC (111)	Та	(Na, Ba)	-1.07
TaC (111)	Та	(Mg, Ca)	-0.73

Surface	On top atoms	Promoters	Adsorption Energy (eV)
TiC (111)	Ti	(Cs, Be)	-0.81
ZrC (111)	Zr	(Cs, Sr)	-1.07
ZrC (111)	Zr	(Li, Cs)	-1.02

**Table 3.7** Adsorption energies of different combination of transition metal carbides and promoter for (111) surfaces in which carbon is on top containing fcc hollow site.

Surface	On top atoms	Promoters	Adsorption Energy (eV)
HfC (111)	С	(Na, Ba)	-1.95
HfC (111)	С	(Na, Na)	-2.02
CrC (111)	С	(K, Sr)	-1.34
NbC (111)	С	(Be, Ba)	-1.14
NbC (111)	С	(K, Ba)	-1.20
MoC (111)	С	(Rb, Be)	-1.17
VC (111)	С	(Li, K)	-1.25
VC (111)	С	(Cs, Mg)	-1.30
WC (111)	W, C	(Rb, Sr)	-0.84
WC (111)	W, C	(Cs, Cs)	-0.51
WC (111)	W, C	(K, Ca)	-0.90
WC (111)	W, C	(Li, Na)	-0.61
TaC (111)	С	(Li, Sr)	-1.53
TiC (111)	С	(Na, Rb)	-1.59
TaC (111)	С	(Be, Sr)	-3.76
TiC (111)	С	(Mg, Sr)	-3.24
Surface	On top atoms	Promoters	Adsorption Energy (eV)
-----------	--------------	-----------	------------------------
ZrC (111)	С	(K. Cs)	-0.98
ZrC (111)	С	(Na, Cs)	-0.89
ZrC (111)	С	(Rb, Ba)	-0.67
ZrC (111)	С	(Ca, Ba)	-1.16

**Table 3.8** Adsorption energies of different combination of transition metalcarbides and promoter for (111) surfaces in which carbon is on topcontaining hcp hollow site.

Surface	On top atoms	Promoters	Adsorption Energy
			(eV)
HfC (111)	С	(Na, Ba)	-1.9
HfC (111)	С	(Na, Na)	-1.71
CrC (111)	С	(K, Sr)	-2.48
NbC (111)	С	(Be, Ba)	-1.34
NbC (111)	С	(K, Ba)	-1.58
VC (111)	С	(Li, K)	-1.3
VC (111)	С	(Cs, Mg)	-1.14
WC (111)	W, C	(Rb, Sr)	-0.84
WC (111)	W, C	(Cs, Cs)	-0.51
WC (111)	W, C	(K, Ca)	-0.90
WC (111)	W, C	(Li, Na)	-0.61
TiC (111)	С	(Na, Rb)	-1.38
TaC (111)	С	(Be, Sr)	-1.72
TiC (111)	С	(Mg, Sr)	-1.55
ZrC (111)	С	(K. Cs)	0.65

Surface	On top atoms	Promoters	Adsorption Energy
			(eV)
ZrC (111)	С	(Na, Cs)	0.47
ZrC (111)	С	(Rb, Ba)	0.46
ZrC (111)	С	(Ca, Ba)	0.40

## Conclusion

We have developed a methodology to identify active transition metal carbide catalysts for the HER from a set of nine considered carbides (TiC, ZrC, HfC, VC, NbC, TaC, CrC, MoC, WC). Our approach incorporates the promoter effect by combining transition metal carbides with alkali and alkaline earth metals as promoters to determine the optimal catalystpromoter combinations for the HER. Initially, we planned to analyze 81 combinations from a total of 1485 possibilities. However, due to the timeconsuming nature of the calculations, we were only able to compute DFT results for 60 combinations, which served as the training dataset for the supervised ML models. Among the ML models employed, the Kernel Ridge Regression (KRR) models demonstrated accurate predictions of the adsorption energy for both the datasets. Both the datasets contain combined (111) and (100) surfaces whereas, one of them contain adsorption energy data on (111) surface at hcp-hollow sites and for other cases, adsorption takes place at the fcc-hollow sites of (111) surfaces. All (100) surfaces only have ffh sites. RMSE and MAE values for the datasets having fcc and ffh sites from KRR are 0.34eV and 0.27eV repectively and for datasets having hcp and ffh sites are 0.35eV and 0.29eV. Among these predictions, the prediction accuracy of the dataset having adsorption energy on fcc and ffh sites, is slightly better. Although we know that this much accuracy is not enough to screen the active catalysts. Hence, in future, we are planning to use active learning technique to reduce the cost of producing training data and to achieve better accuracy. It also can help us to choose the datapoints with more uncertainty which may subsequently help in the process of the prediction for unknown search space of the catalysts. For this, we will utilize the DFT results of Pt as a reference to screen the catalysts. By minimizing the reliance on extensive DFT calculations and leveraging intelligent ML algorithms, our approach significantly expedites the discovery of heterogeneous catalysts We envision that this methodology

will facilitate the rapid design of various other heterogeneous catalysts within a shorter timeframe.

## REFERENCES

1. Pandit, N. K.; Roy, D.; Mandal, S. C.; Pathak, B. Rational designing of bimetallic/trimetallic hydrogen evolution reaction catalysts using supervised machine learning. *J. Phys. Chem. Lett.* **2022**, *13*, *32*, 7583–7593 (DOI: 10.1021/acs.jpclett.2c01401)

2. Chen, W. F.; Muckerman, J. T.; Fujita, E. Recent developments in transition metal carbides and nitrides as hydrogen evolution electrocatalysts. *Chem. Commun.*, **2013**, *49*, 8896-8909 (DOI: 10.1039/C3CC44076A)

3) Debe, M. K. Electrocatalyst approaches and challenges for automotive fuel cells. *Nature* **2012**, *486*, 43–51. (DOI: 10.1038/nature11115)

4) Zheng, Y.; Jiao, Y.; Jaroniec, M.; Qiao, S. Z. Advancing the electrochemistry of the hydrogen-evolution reaction through combining experiment and theory. *Angew. Chem. Int.* Ed. **2015**, *54*, 52–65. (DOI: 10.1002/anie.201407031)

5) Tiwari, J. N.; Dang, N. K.; Sultan, S.; Thangavel, P.; Jeong, H. Y.; Kim, K. S.; Multi-heteroatom-doped carbon from waste-yeast biomass for sustained water splitting. *Nat Sustain* **2020**, *3*, 556–563 (DOI: 10.1038/s41893-020-0509-6)

6) Chialvo, M. R. Gennero.; Chialvo, A. C. Kinetics of hydrogen evolution reaction with frumkin adsorption: re-examination of the volmer-heyrovsky and volmer-tafel routes *Electrochimica Acta* **1998**, *44*, 841-851 (DOI: 10.1016/S0013-4686(98)00233-3)

7) Sabatier, P. Hydrogeneous et Deshydrogenations par Catalyse. Ber. Dtsch. Chem. Ges. **1911**, *44*, 1984-2001 (DOI: 10.1002/cber.19110440303)

 Beng, J.; Deng, D.; Bao, X. Robust Catalysis on 2D materials encapsulating metals: Concept, Application, and Perspective. Adv. Mater.
2017, 1606967 (DOI: 10.1002/adma.201606967)

9) Li, S., Chen, B., Wang, Y. et al. Oxygen-evolving catalytic atoms on metal carbides. Nat. Mater. **2021** *20*, 1240–1247 (DOI: 10.1038/s41563-021-01006-2)

10) Eames, C.; Islam, M. S. Ion intercalation into two-dimensional transition-metal carbides: global Screening for new high-capacity battery materials. J. Am. Chem. Soc. **2014**, *136*, *46*, 16270–16276 (DOI: 10.1021/ja508154e)

11) Fischer, A.; Muller, J. O.; Antonietti, M.; Thomas, A. synthesis of ternary metal nitride nanoparticles using mesoporous carbon nitride as reactive template ACS Nano **2008**, *2*, *12*, 2489–2496 (DOI: 10.1021/nn800503a)

12) Tiwari, J. N.; Harzandi, A. M.; Ha, M.et. al. High-performance hydrogen evolution by Ru single atoms and nitrided-Ru nanoparticles implanted on N-Doped graphitic sheet. Adv. Energy Mater.**2019**, *9*, 1900931 (DOI: 10.1002/aenm.201900931)

13) Stevens, M. B.; Kreider, M. E.; Patel, A. M. Nitride or oxynitride? elucidating the composition–activity relationships in molybdenum Nitride Electrocatalysts for the Oxygen Reduction Reaction. Chem. Mater. **2020**, *32*, *7*, 2946–2960 (DOI: 10.1021/acs.chemmater.9b05212)

14) Umer, M.; Umer S.; Zafari M.; Ha, M.; Anand, R.; Hajibabaei, A.; Abbas, A.; Lee, G.; Kim, K, S. Machine learning assisted high-throughput screening of transition metal single atom based superb hydrogen evolution electrocatalysts. J. Mater. Chem. A, **2022**, *10*, 6679–6689 (DOI: 10.1039/D1TA09878K)

15) Johansson, L. I. Electronic and structural properties of transition-metal carbide and nitride surfaces. *Surf. Sci. Rep.* **1995**, *21*, 177 (DOI: 10.1016/0167-5729(94)00005-0)

16) Oyama, S. T. The chemistry of transition metal carbides and nitrides; blackie academic and professional: Glasgow, **1996**. (DOI: 10.1007/978-94-009-1565-7\_1)

17) Chen, J. G. Carbide and nitride overlayers on early transition metal surfaces: preparation, characterization, and reactivities. Chem. Rev. **1996**, *96*, 1497 (DOI: 10.1021/cr950232u)

18) Chen, J. G. NEXAFS investigations of transition metal oxides, nitrides, carbides, sulfides and other interstitial compounds. *Surf. Sci. Rep.* 1997, *30*, 1 (DOI: 10.1016/S0167-5729(97)00011-3)

19) Levy, R. L.; Boudart, M. Science **1973**, *181*, 547 (DOI: 10.1126/science.181.4099.547)

20) Hwu, H. H.; Chen, J, G. Surface chemistry of transition Metal carbides. *Chem. Rev.* **2005**, *105*, 185–212 (DOI: 10.1021/cr0204606)

21) Chen, J, G. Carbide and nitride overlayers on early transition metal surfaces: preparation, characterization, and reactivities. *Chem. Rev.* **1996**, 96, 1477–1498 (DOI: 10.1021/cr950232u)

22) Siegel, E. D-Bandwidth Contraction upon Metalloid formation: A resolution of the bonding controversy in transition metal carbides, nitrides, and borides. semicond. insul. **1979**, *5*, 47

23) Heine, V. s-d Interaction in Transition Metal. *Phys. Rev.* **1967**, 153, 673 (DOI: 10.1103/PhysRev.153.673)

24) Strmcnik, D.; Kodama, K.; Vliet, D.; Greeley, J.; Stamenkovic, V, R.; Markovic, N, M. The role of non-covalent interactions in electrocatalytic fuel-cell reactions on platinum. *Nature chemistry* **2009**, 466-472 (DOI: 10.1038/nchem.330)

25) Colic, V.; Pohl, M, D.; Scieszka, D.; Bandarenka, A, S. Influence of the electrolyte composition on the activity and selectivity of electrocatalytic centers. *Catalysis Today* **2015** (DOI: 10.1016/j.cattod.2015.08.003)

26). Tokuoka, M. Polarograpic studies with the dropping mercury kathode. Part XXVII. The electro-reduction and estimation of nitrates and nitrites. *Collect. Czech. Chem. Commun.* **1932**,*4*, 444-455 (DOI: 10.1135/cccc19320444)

27) Lee, W.; Han, J. W.; Chen, Y.; Cai, Z.; Yildiz, B. Cation size mismatch and charge interactions drive dopant segregation at the surfaces of manganite perovskites. *J. Am. Chem. Soc.* **2013**, *135*, 7909–7925 (DOI: 10.1021/ja3125349)

28) Stoffelsma, C.; Rodriguez, P.; Garcia, G.; Araez, N. G.; Strmcnik, D.; Markovic, N. M.; Koper, M. T. M. Promotion of the oxidation of carbon monoxide at stepped platinum single-crystal electrodes in alkaline media by lithium and beryllium cations. *J. AM. HEM. SOC.* **2010**, *132*, 16127–16133 (DOI: 10.1021/ja106389k)

29) Xue, S.; Garlyyev, B.; Watzele, S.; Liang, Y.; Fichtner, J.; Pohl M. D.; and Bandarenka, A. S. Influence of alkali metal cations on the hydrogen evolution reaction activity of Pt, Ir, Au and Ag electrodes in alkaline electrolytes. *ChemElectroChem.* **2018**, 2326–2329 (DOI: 10.1002/celc.20180069030).

30) Turing, A. M. Computing Machinery and Intelligence, Mind, LIX,1950, 236, 433-460

31) Pedersen, J. K.; Batchelor, T. A. A.; Bagger, A.; Rossmeisl, J. Highentropy alloys as catalysts for the CO<sub>2</sub> and CO reduction reactions. *ACS Catal.* **2020**, *10*, 2169–2176 (DOI: 10.1021/acscatal.9b04343) 32. Roy, D.; Mandal, S. C.; Pathak, B. Machine Learning-Driven High-Throughput Screening of Alloy-Based Catalysts for Selective CO<sub>2</sub> Hydrogenation to Methanol. *ACS Appl. Mater. Interfaces.* **2021**, *13*, *47*, 56151-56163 (DOI: 10.1021/acsami.1c16696)

33. Wu, S.; Wang, Z.; Zhang, H.; Cai, J.; Li, J. Deep Learning Accelerates the Discovery of Two-Dimensional Catalysts for Hydrogen Evolution Reaction. Energy Environ. Mater. **2022** (DOI: 10.1007/s12274-022-4265y)

34. Baghban, A.; Habibzadeh, S.; Zokaee A., F. On the evaluation of hydrogen evolution reaction performance of metal-nitrogen-doped carbon electrocatalysts using machine learning technique. Sci Rep **2021**, *11*, 21911 (DOI: 10.1038/s41598-021-00031-0)

35. Mao, X.; Wang, L.; Xu, Y.; Wang, P.; Li, Y.; Zhao, J. Computational high-throughput screening of alloy nanoclusters for electrocatalytic hydrogen evolution. *npj Comput Mater*, **2021**, *7*, 46 (DOI: 10.1038/s41524-021-00514-8)

36. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 2011, 12, 2825–2830

37. Xia B. Y., Wu H.B., Wang X., Lou X.W. Index facets and enhanced electrocatalytic properties, *Angew. Chem. Int. Ed*, **2013**, *52*, 12337-12340 (DOI: 10.1002/ange.201307518)

38. Mortensen, J. J.; Hansen, L. B.; Jacobsen, K. W. Real-Space Grid Implementation of the Projector Augmented Wave Method. Phys. Rev. B: Condens. Matter Mater. Phys. **2005**, *71*, 035109 (DOI: 10.1103/PhysRevB.71.035109) 39. Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dułak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. A.; Kristoffersen, H. H.; Kuisma, M.; Larsen, A. H.; Lehtovaara, L.; Ljungberg, M.; Lopez-Acevedo, O.; Moses, P. G.; Ojanen, J.; Olsen, T.; Petzold, V.; Romero, N. A.; Stausholm-Møller, J.; Strange, M.; Tritsaris, G. A.; Vanin, M.; Walter, M.; Hammer, B.; Häkkinen, H.; Madsen, G. K. H.; Nieminen, R. M.; Nørskov, J. K.; Puska, M.; Rantala, T. T.; Schiøtz, J.; Thygesen, K. S.; Jacobsen, K. W. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys.: Condens. Matter.* **2010**, *22*, 253202 (DOI: 10.1088/0953-8984/22/25/253202)

40. Perdew J.P., Burke K., Ernzerhof M. Generalized gradient approximation made simple, *Phys. Rev. Lett.* **1996**, *77*, 3865-3868 (DOI: 10.1103/PhysRevLett.77.3865)

41. Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **1997**, 78, 1396 (DOI: 10.1103/PhysRevLett.77.3865)