UNDERSTANDING REACTIVITY DESCRIPTORS IN AB INITIO DESIGN OF A HETEROGENEOUS CATALYST USING MACHINE LEARNING APPROACHES

AMRISH KUMAR

DEPARTMENT OF CHEMICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI

MAY 2024

© Indian Institute of Technology Delhi (IITD), New Delhi, 2024

UNDERSTANDING REACTIVITY DESCRIPTORS IN AB INITIO DESIGN OF A HETEROGENEOUS CATALYST USING MACHINE LEARNING APPROACHES

by

AMRISH KUMAR

DEPARTMENT OF CHEMICAL ENGINEERING

Submitted

In fulfilment of the requirements of the degree of doctor of philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI
MAY 2024

I dedicate this thesis to my parents, family, and friends

Certificate

This is to certify that the thesis entitled **"Understanding reactivity descriptors in ab initio design of a heterogeneous catalyst using machine learning approaches"** submitted by Mr. Amrish Kumar to the Indian Institute of Technology Delhi for the award of the degree of Doctor of Philosophy, is a record of the original bonafide research work carried out by him. He has worked under my supervision and has fulfilled the requirements, which to my knowledge, has reached the requisite standard for the submission of this thesis. The results contained in this thesis have not been submitted in part or full to any University or Institute for the award of any degree or diploma.

Prof. Manojkumar C Ramteke	**Prof. M Ali Haider**
Supervisor	Co-Supervisor
Department of Chemical Engineering	Department of Chemical Engineering
Indian Institute of Technology Delhi	Indian Institute of Technology Delhi

Acknowledgements

Every thesis project is the result of the knowledge and experience gained by the student conducting the corresponding research. However, no project is possible without proper guidance and help extended to the students by experts in the field, with the sole benevolent purpose of contributing to the intellectual development of the students. I take this opportunity to express my deep sense of gratitude and respect towards **Prof. Manojkumar C Ramteke** and **Prof. M Ali Haider**, who provided me their guidance and support for me to complete this work. I am grateful to them for their invaluable guidance, supervision, encouragement, and constructive criticism, without which it would not have been possible for me to deliver this thesis work.

I am also thankful to the Head of the Department of Chemical Engineering at the Indian Institute of Technology Delhi, and all the faculty members of the department for their support and valuable suggestions. I wish to convey my gratitude to members of my research committee Prof. K K Pant, Prof. Shantanu Roy and Prof. B K Panigrahi for their valuable suggestions and critical review of my work.

I express my special thanks to my laboratory colleagues: Dr. Sonit Balyan, Dr. Deepak Jangra, Dr. Tuhin Khan, Dr. Uzma Anjum, Dr. Shelakha Gupta, Dr. Debashish Panda, Dr. Feleke Bayu, Dr. Sourabh Mishra, Dr. Gourav, Mr. Jayendran Iyer, Mr. M Ussama, Mr. Deepak Seth, Mr. Shantanu Sontakke, Mr. Shivam Saxena, Mr. Deepak Singh, Ms. Anubha Aggarwal, Ms. Mili Bang, Ms. Aradhna Singh, Ms. Chaitra Shenoy, Ms. K V Haseena, Ms. Iqra, Mr. Asutosh Negi, Mr. Mahdiyar, Mr. Gulshitab Alam, and Mr. Adarsh Madhu for their valuable suggestions and camaraderie.

I am thankful to all my friends of IIT Delhi: Dr. Rajan Singh, Dr. Devendra Kumar, Dr. Anu Kadyan, Dr. Vandit Vijay, Dr. Joyjit Mukherjee, Dr. Sanchyan Pal, Dr. Prashant Jadhao, Ms. Vallari, Dr. Nalin Sharma, Dr. Shabina Ashraf, Dr. Neelam Choudhary, Dr. Khyati Verma, Dr. Uma Dwivedi, Dr. Supriyo Chakraborty, Mr. Vaibhav Pandey, Mr. Puneet Sharma, Dr. Sourav Patranabish, Dr. Sumallya

Mukhopadhyay, Dr. Preetam, Mr. Devashish Phadnis, Mr. Rahul Shubham, Mr. Manoj Kumar and Ms Monika Rao.

Last but not least, I thank my parents Mr. Prem Das and Mrs. Meenu Das for their blessings and love. I would like to thank both of my brothers Mr. Dushyant Kumar and Dr. Rohitash Kumar for their direct or indirect support in the completion of my work.

At last, I thank the Almighty God for his blessing and giving me patience, strength, and peace throughout the journey of life.

Amrish Kumar

Abstract

In this work, combined Machine learning (ML) and Density Functional Theory (DFT) approaches were used in order to find descriptors of the catalytic activity of various alloys. In heterogeneous catalysis, binding energy of an adsorbate is an important descriptor for the estimation of catalytic activity. Therefore, DFT calculations were first conducted to compute the binding energies of atomic oxygen (E_O) and carbon (E_C) on various facets [(111), (110), (100)] of late transition metals. In this initial study, we investigated a ML approach to predict these energies. Gradient Boosting Regression (GBR) emerged as the most effective model, achieving low errors (RMSE of 0.14 eV for E_O and 0.31 eV for E_C) while considering facet-dependent adsorption.

Building upon this success of ML-DFT approach, the second study utilized DFT calculated E_O and E_C over group 11 based single atom alloys (SAAs), which were used as the target variable. Properties of the host (Cu, Ag, and Au) and guest (dopants) metals were used as input features to construct a dataset for building a ML model. The GBR model was selected after a comparative study to predict binding energies on SAAs, as it yields the lowest test errors of 0.22 eV and 0.19 eV for E_O and E_C, respectively. Later, this ML approach was combined with microkinetic modelling (MKM) for non-oxidative dehydrogenation of ethanol (NODH). Turnover frequencies for ethanol conversion on NiAu, NiAg, and PtAg on SAAs were predicted to be increased four-fold (10^{-1} s^{-1} and 10^{-3} s^{-1}) compared to their respective monometallic counterparts (10^{-5} s^{-1} and 10^{-7} s^{-1}), which is of the same order of magnitude derived from DFT.

Furthermore, for the NODH reaction of ethanol to produce acetaldehyde, the third study extended the aforementioned SAA studies to Ni, Pd, and Pt-doped Cu based diatomic (DAAs) and triatomic (TAAs) alloys and compared the O-H and α-C-H activation barriers obtained in each case. Two distinct routes over DAAs and TAAs have been investigated. Route 1, involving O-H activation followed by α-C-H activation, shows improved O-H activation barrier from 1.03 eV to 0.69 eV over Ni doped SAAs and TAAs, respectively. In contrast, no significant O-H barrier reduction is observed for Ni doped DAAs (1.04 eV) compared to SAAs (1.03 eV). Route 2, where α-C-H activation occurs first, exhibits a lower α-C-H bond

cleavage barrier in both Ni doped DAAs (0.83 eV) and TAAs (0.59 eV) compared to the O-H scission in Route 1. Additionally, increasing concentrations of Pd and Pt on the Cu surface do not lead to significant barrier reductions in O-H and α-C-H activation, suggesting the need for further exploration of active sites.

Subsequently, in our fourth study, the non-oxidative coupling of methane (NOCM) on a series of transition metal doped graphene based dual metal site catalysts (DMSCs) was investigated using ML and DFT. The DFT computed adsorption energies and transition state energies for the initial C-H activation step of NOCM are used as target variables for training the ML models. In addition, several readily available features (viz. atomic number, atomic mass, group, period, radius, electronegativity, melting point, boiling point, heat of fusion, ionization energy, surface energy, density, d-band center, d-band filling, and Wigner-Seitz radius) associated with both the transition metals doped on graphene are used as input features; which are then utilized by ML models to predict the catalytic activity of DMSCs. In this study, the extra tree regression (ETR) model emerges as the best performer, accurately predicting transition state and adsorption energies with RMSEs of 0.27 and 0.33 eV, respectively.

From the aforementioned studies, the ML approach is demonstrated to provide a rationale for feature selection from periodic properties of the elements to synthesize catalytically active SAAs and DMSCs for high throughput in-silico screening.

सार

इस कार्य में, विभिन्न मिश्र धातुओं की उत्प्रेरक गतिविधि के विवरणक खोजने के लिए मशीन लर्निंग (ML) और डेंसिटी फंक्शनल थ्योरी (DFT) दृष्टिकोणों को मिलाया गया है। विषमजीन उत्प्रेरण में, एक एडसॉर्बेट की बाइंडिंग ऊर्जा एक महत्वपूर्ण विवरणक है जो उत्प्रेरक गतिविधि के अनुमान के लिए महत्वपूर्ण होती है। इसलिए, DFT गणनाओं का उपयोग करके पहले विभिन्न लेट ट्रांज़िशन धातुओं के विभिन्न पहलुओं [(111), (110), (100)] पर परमाणु ऑक्सीजन (E_O) और कार्बन (E_C) की बाइंडिंग ऊर्जा की गणना की गई। इस प्रारंभिक अध्ययन में, हमने इन ऊर्जाओं की भविष्यवाणी करने के लिए एक मशीन लर्निंग दृष्टिकोण का परीक्षण किया। ग्रेडिएंट बूस्टिंग रिग्रेशन (GBR) सबसे प्रभावी मॉडल के रूप में उभरा, जिसमें पहलू-निर्भर विज्ञापन पर विचार करते समय E_O के लिए 0.14 eV और E_C के लिए 0.31 eV की कम त्रुटियों (RMSE) प्राप्त हुईं।

ML-DFT दृष्टिकोण की इस सफलता के आधार पर, दूसरे अध्ययन में समूह 11 आधारित सिंगल एटम एलॉय (SAAs) पर DFT गणना E_O और E_C का उपयोग किया गया, जिन्हें लक्ष्य चर के रूप में उपयोग किया गया। होस्ट (Cu, Ag, और Au) और गेस्ट (डोपेंट्स) धातुओं के गुणों का उपयोग इनपुट फीचर्स के रूप में किया गया ताकि डेटासेट का निर्माण किया जा सके और एक मशीन लर्निंग मॉडल बनाया जा सके। GBR मॉडल को चयनित किया गया, क्योंकि इसने E_O और E_C के लिए क्रमशः 0.22 eV और 0.19 eV की सबसे कम परीक्षण त्रुटियों के साथ बाइंडिंग ऊर्जा की भविष्यवाणी की। बाद में, इस मशीन लर्निंग दृष्टिकोण को माइक्रोकिनेटिक मॉडलिंग (MKM) के साथ मिलाया गया, ताकि एथेनॉल के गैर-ऑक्सीकरण डीहाइड्रोजनेशन (NODH) की टर्नओवर फ्रीक्वेंसी की भविष्यवाणी की जा सके। NiAu, NiAg, और PtAg SAAs पर एथेनॉल रूपांतरण की टर्नओवर फ्रीक्वेंसी को उनके संबंधित मोनोमेटलिक समकक्षों की तुलना में चार गुना बढ़ाकर (10^{-1} s^{-1} और 10^{-3} s^{-1}) (10^{-5} s^{-1} और 10^{-7} s^{-1}) प्राप्त किया गया, जो DFT से प्राप्त की गई उसी क्रम की परिमाण की है।

इसके अलावा, एथेनॉल से एसीटैल्डिहाइड के उत्पादन के लिए NODH प्रतिक्रिया के लिए, तीसरे अध्ययन में उपरोक्त SAA अध्ययनों को Ni, Pd, और Pt-डोप्ड Cu आधारित डायटॉमिक (DAAs) और ट्रायटॉमिक (TAAs) मिश्र धातुओं तक बढ़ाया गया और प्रत्येक मामले में O-H और α-C-H सक्रियण बाधाओं की तुलना की गई। DAAs और TAAs पर दो विशिष्ट मार्गों की जांच की गई। मार्ग 1, जिसमें O-H सक्रियण पहले होता है और उसके बाद α-C-H सक्रियण होता है, दिखाता है कि Ni डोप्ड SAAs और TAAs पर O-H सक्रियण बाधा 1.03 eV से 0.69 eV तक सुधारित होती है। इसके विपरीत, Ni डोप्ड DAAs (1.04 eV) की तुलना में SAAs (1.03 eV) के लिए O-H बाधा में कोई महत्वपूर्ण कमी नहीं देखी जाती है। मार्ग 2, जिसमें α-C-

H सक्रियण पहले होता है, Ni डोप्ड DAAs (0.83 eV) और TAAs (0.59 eV) में O-H स्कीशन की तुलना में α-C-H बंधन के विखंडन बाधा को कम प्रदर्शित करता है। इसके अलावा, Cu सतह पर Pd और Pt की बढ़ती सांद्रता O-H और α-C-H सक्रियण में महत्वपूर्ण बाधा कमी नहीं दिखाती है, जिससे सक्रिय स्थलों की और अधिक खोज की आवश्यकता का सुझाव मिलता है।

इसके बाद, चौथे अध्ययन में, ट्रांज़िशन मेटल डोप्ड ग्रेफीन आधारित डुअल मेटल साइट उत्प्रेरकों (DMSCs) पर मीथेन के गैर-ऑक्सीकरण युग्मन (NOCM) की जांच मशीन लर्निंग और DFT का उपयोग करके की गई। NOCM के प्रारंभिक C-H सक्रियण चरण की DFT गणना एडसॉर्पशन ऊर्जा और संक्रमण स्थिति ऊर्जा का उपयोग लक्ष्य चर के रूप में किया गया है, जो ML मॉडलों के प्रशिक्षण के लिए उपयोग किया गया। इसके अतिरिक्त, कई उपलब्ध फीचर्स (जैसे परमाणु संख्या, परमाणु द्रव्यमान, समूह, अवधि, त्रिज्या, विद्युत-ऋणात्मकता, गलनांक, क्वथनांक, संलयन की ऊष्मा, आयनीकरण ऊर्जा, सतह ऊर्जा, घनत्व, डी-बैंड केंद्र, डी-बैंड भरना, और विग्नर-सीट्ज त्रिज्या) ट्रांज़िशन मेटल्स के साथ ग्रेफीन पर डोप्ड के जुड़े फीचर्स इनपुट के रूप में उपयोग किए गए हैं; जिन्हें फिर ML मॉडल्स द्वारा DMSCs की उत्प्रेरक गतिविधि की भविष्यवाणी के लिए उपयोग किया जाता है। इस अध्ययन में, एक्स्ट्रा ट्री रिग्रेशन (ETR) मॉडल सबसे अच्छा प्रदर्शनकर्ता बनकर उभरा, जिसने क्रमशः 0.27 और 0.33 eV के RMSEs के साथ संक्रमण स्थिति और एडसॉर्पशन ऊर्जा की सही भविष्यवाणी की।

उपरोक्त अध्ययनों से, ML दृष्टिकोण को तत्वों के आवर्त गुणों से विवरणक चयन के लिए एक तर्क प्रदान करने के लिए प्रदर्शित किया गया है ताकि उच्च थ्रूपुट इन-सिलिको स्क्रीनिंग के लिए उत्प्रेरक रूप से सक्रिय SAAs और DMSCs का संश्लेषण किया जा सके।

Table of Contents

Certificate i

Acknowledgements ii

Abstract iv

List of Figures x

List of Tables xiv

Chapter 1 : Introduction 1

- 1.1 Background and Motivation 1
- 1.2 Literature Survey 5
  - 1.2.1 Dilute Metal Alloys and Catalysts 5
  - 1.2.2 ML Based Reactivity Descriptor Search 9
- 1.3 Flow of Thesis 19
- 1.4 Goals and Objectives 22
- 1.5 Thesis Organization 22

Chapter 2 : Computational Methodology 23

- 2.1 Density Functional Theory Calculations 23
- 2.2 Machine Learning 30
  - 2.2.1 K-Nearest Neighbour 30
  - 2.2.2 Support Vector Regression 31
  - 2.2.3 Random Forest and Extra Tree Regression 33
  - 2.2.4 Gradient Boosting Regression 34
  - 2.2.5 Artificial Neural Network 35
- 2.3 Microkinetic Modeling 37

Chapter 3 : Machine Learning Enabled Prediction of Descriptors for Catalyst Reactivity 39

- 3.1 Results and Discussion 39
- 3.2 Summary 47

Chapter 4 : Integrated Machine Learning and Microkinetic Modeling for Accelerated Screening of Dilute Metal Alloy Catalysts 49

- 4.1 Results 49
  - 4.1.1 Binding Energy Trends on SAAs 49
  - 4.1.2 Feature Selection 52
  - 4.1.3 Gradient Boosting Regression 57
  - 4.1.4 Microkinetic Modeling 60
- 4.2 Discussion 62

4.3 Summary 64
Chapter 5 : Mechanistic Insights into the Reactivity of Dilute Metal Alloy Catalysts 66
5.1 Results and Discussion 66
5.2 Summary 74
Chapter 6 : A Machine Learning Approach to Predict Activation Energies 75
6.1 Results and Discussion 75
6.2 Summary 92
Chapter 7 : Conclusions and Future Scope 93
References 95
Appendix A 105
Appendix B 114

List of Figures

Figure 1.1. Evolution of heterogeneous catalysis: From Paul Sabatier's principle to machine learning...... 2

Figure 1.2. Catalyst discovery through data analysis (employing machine learning algorithms) and quantum mechanical (utilizing available materials databases) approaches. 4

Figure 1.3. Surfaces explored in this study for non-oxidative dehydrogenation (NODH) of ethanol and non-oxidative coupling of methane (NOCM) are **(a)** single atom alloys, **(b)** diluted alloys, and **(c)** dual metal site catalysts. **Orange:** coinage metals (Cu, Ag, and Au), **Light blue:** transtion metals, **Dark blue:** nitrogen, **Grey:** carbon. 7

Figure 1.4. Descriptors in catalysis to screen wide range of catalyst surfaces are **(a)** adsorption energy, **(b)** activation barrier and d band center, **(c)** transition state energy, and **(d)** intermediate energy. This is a representative plot of catalytic descriptors used in all the chapters. 11

Figure 2.1. Self-consistent iterative method to solve Kohn-Sham equations... 24

Figure 2.2. Surface sites are illustrated on **(a)** (100), **(b)** (110) and **(c)** (111). **Dark grey:** top layer, **Light grey:** bottom layer.. 28

Figure 2.3. A schematic of building a dataset for machine learning models to predict catalytic reactivity descriptors for SAAs (single atom alloys). 29

Figure 2.4. Representation of K nearest neighbour (KNN) model using **(a)** Euclidean Distance for **(b)** Prediction of new data point. This diagram illustrates the KNN algorithm used to forecast the classification of a new data point (in green). The Euclidean distance is employed to determine the K nearest neighbors using the training data consisting of both red and blue mark. In a classification task, the new point is assigned the class label that appears most frequently (red in this case) among its K nearest neighbors. ... 31

Figure 2.5. Representation of support vector regression (SVR) model. The data points, in red, represent the training data. The dotted line area represents the epsilon-tube, which is a region that allows for a certain level of prediction errors. The solid line represents the hyperplane of the SVR model. The purpose of the hyperplane is to fit within the tube and maximize the distance to the most distant data points, which are referred to as support vectors. New data points will be predicted by considering their position in relation to the hyperplane. 32

Figure 2.6. Representation of random forest (RFR) and extra tree (ETR) regression models. The ensemble comprises of decision trees, each grown on a random subset of data and using a random feature set at each split. The final prediction of the RFR and ETR model is the average of the predictions from all the trees in the ensemble.. 34

Figure 2.7. Representation of the gradient boosting regression (GBR) model. 35

Figure 2.8. Representation of artificial neural network (ANN) model. 36

Figure 2.9. Workflow of machine learning approach. 37

Figure 4.1. DFT calculated binding energy of atomic oxygen E_O (blue), and atomic carbon E_C (red) values of SAAs (using energies of atomic oxygen and atomic carbon calculated in vacuum, from VASP) based on **(a)** Cu, **(b)** Ag and **(c)** Au. Error bar = 0.2 eV. 51

Figure 4.2. Atop oxygen binding configurations on Cu, Ag, and Au based SAAs. Numbers represent distance (Å) between O and dopant single atom. 52

Figure 4.3. Correlation plot for E_O and E_C with the features of **(a)** dopant atoms and **(b)** host metals. The heatmap shows positive or negative correlation of features with E_O and E_C. 53

Figure 4.4. Average RMSE errors for 100 trials are compared for **(a)** E_O and **(b)** E_C, and **(c)** a schematic depiction of GBR workflow. 56

Figure 4.5. Feature importance of dopant single atom for the GBR model for predicting **(a)** E_O and **(b)** E_C of Cu, Ag, and Au based SAAs. 58

Figure 4.6. The d-band scaling plots **(a)** E_O (binding energy of atomic oxygen) versus average d-band center of three atoms (for 3-fold oxygen adsorption) and single atom d-band center (for atop oxygen adsorption), **(b)** E_C (binding energy of atomic carbon) versus average d-band center of three atoms (for 3-fold carbon adsorption). Error bar = 0.2 eV. 59

Figure 4.7. Deviation of **(a)** E_O and **(b)** E_C predicted values from DFT calculated values. The grey points represent training datapoints and the points highlighted in blue and red are test set for oxygen and carbon adsorption, respectively. 61

Figure 4.8. **(a)** Reaction scheme for ethanol NODH via two pathways. **(b)** Comparison of TOFs of ethanol consumption over NiAu, NiAg, and PtAg as obtained from DFT (blue) and ML (green) as plotted from the MKM developed over (111) sites of transition metal catalysts at the reaction conditions of T = 450K and P = 1 bar. Error bar = 0.2 eV. 62

Figure 5.1. Reaction scheme for ethanol NODH via two pathways. 67

Figure 5.2. Reaction Diagram of NODH on Ni_2Cu (111) surface: Comparison of Routes 1 (Black) and 2 (Red). Activation barrier of intrinsic step is indicated on the bar, marking the TS point. Surface-adsorbed species marked with an asterisk (*). Energies are determined in relation to the energy of the ethanol vapour phase. Error bar = 0.2 eV. 68

Figure 5.3. Reaction Diagram of NODH on Ni_3Cu (111) surface: Comparison of Routes 1 (Black) and 2 (Red). Activation barrier of intrinsic step is indicated on the bar, marking the TS point. Surface-adsorbed species marked with an asterisk (*). Energies are determined in relation to the energy of the ethanol vapour phase. Error bar = 0.2 eV. 69

Figure 5.4. Reaction Diagram of NODH on Pd_2Cu (111) surface: Comparison of Routes 1 (Black) and 2 (Red). Activation barrier of intrinsic step is indicated on the bar, marking the TS point. Surface-adsorbed

species marked with an asterisk (*). Energies are determined in relation to the energy of the ethanol vapour phase. Error bar = 0.2 eV. 70

Figure 5.5. Reaction Diagram of NODH on Pd_3Cu (111) surface: Comparison of Routes 1 (Black) and 2 (Red). Activation barrier of intrinsic step is indicated on the bar, marking the TS point. Surface-adsorbed species marked with an asterisk (*). Energies are determined in relation to the energy of the ethanol vapour phase. Error bar = 0.2 eV. 71

Figure 5.6. Reaction Diagram of NODH on Pt_2Cu (111) surface: Comparison of Routes 1 (Black) and 2 (Red). Activation barrier of intrinsic step is indicated on the bar, marking the TS point. Surface-adsorbed species marked with an asterisk (*). Energies are determined in relation to the energy of the ethanol vapour phase. Error bar = 0.2 eV. 72

Figure 5.7. Reaction Diagram of NODH on Pt_3Cu (111) surface: Comparison of Routes 1 (Black) and 2 (Red). Activation barrier of intrinsic step is indicated on the bar, marking the TS point. Surface-adsorbed species marked with an asterisk (*). Energies are determined in relation to the energy of the ethanol vapour phase. Error bar = 0.2 eV. 73

Figure 6.1. Transition metals selected from the periodic table to study different DMSCs (Dual Metal Site Catalysts) configurations for methane activation. DFT calculations performed for methane activations on the combination of metals highlighted in red (M1) and green (M2). The product state structure obtained on methane activation shows the bond distances as dashed lines. 76

Figure 6.2. C-H bond activation energy trend for methane on DMSCs (dual metal site catalysts) built using combinations of M1: **(a)** Cr, **(b)** Mn, and **(c)** Fe with the M2 transition metals (Co, Ni, and Cu) belonging to the same row and (Co, Rh and Ir) belonging to the same group in the periodic table. Activation barrier of intrinsic step is indicated as the peak of the bar, marking the TS point. All other calculations of different combinations of transition metals are available in Appendix B. Error bar = 0.2 eV. 78

Figure 6.3. Activation energy (E_a) as a function of **(a)** average d-band center ($\boldsymbol{\varepsilon_m}$), **(b)** C-H bond distances in the TS structure and **(c)** M2-H bond distances in the TS structure; for the DMSCs (dual metal site catalysts) configurations shown in Figure 6.2. All the bond distances M1-C, M1-H, M2-H, M1-M2 and C-H (C and H from CH_4) are reported in Appendix B. Error bar = 0.2 eV. 79

Figure 6.4. Workflow for predicting C-H activation on N-coordinated graphene based DMSCs (dual metal site catalysts) using the ML approach, which includes, data acquisition, feature selection, model training, and model testing. 81

Figure 6.5. Pearson correlation plot depicting linear dependency between feature space and target space. 82

Figure 6.6. Comparison of ML model performance based on RMSE for **(a)** Activation Energy E_a, **(b)** Final State Adsorption Energy $E_{CH3*, H*}$. 85

Figure 6.7. Working of **Ext**remely **Ra**ndomized Tree Regression (ETR) also known as Extra Tree Regression Model. 86

Figure 6.8. Representation of DFT computed (solid) and ML predicted (dashed) reaction energetics for the C-H activation from the best performing ETR model. Error bar = 0.2 eV. 87

Figure 6.9. Feature importance for $E_{CH3*, H*}$ prediction using the ETR model. 88

Figure 6.10. Feature importance for E_a prediction using the ETR model. 90

Figure 6.11. Deviation of ML predicted (ETR model) **(a)** E_a, and (b) $E_{CH3*, H*}$ from DFT calculations. ... 91

Figure A.1. The geometries of carbon **(a)**, **(c)**, **(e)** and oxygen **(b)**, **(d)**, **(f)** adsorbed on Cu, Ag, and Au based SAAs, respectively. The numbers represent adsorbate distance (Å) from single atom. 110

Figure A.2. Relative feature importance of host atom in GBR model for predicting **(a)** E_O and **(b)** E_C of Cu, Ag and Au based SAAs 111

Figure A.3. **(a)** E_O versus average d-band center of 3 atoms, **(b)** E_O versus single atom d-band center and **(c)** E_C versus single atom d-band center. Error bar = 0.2 eV. 112

Figure B.1. Bond distances for DMSCs built using combinations of M1 transition metals (Cr, Mn and Fe) with the M2 transition metals (Co, Ni and Cu) belonging to the same row in the periodic table. 117

Figure B.2. Bond distances for DMSCs built using combinations of M1 transition metals (Co, Rh and Ir) with the M2 transition metals (Co, Rh and Ir) belonging to the same group in the periodic table. 119

Figure B.3. PDOS for DMSCs built using combinations of M1 transition metals (Cr, Mn and Fe) with the M2 transition metals (Co, Ni and Cu) belonging to the same row in the periodic table. 120

Figure B.4. C-H bond activation energy trend for methane on DMSCs build using combinations of M1 transition metals (Co, Rh and Ir) with the M2 transition metals (Co, Rh and Ir) belonging to the same group in the periodic table. Error bar = 0.2 eV. 121

Figure B.5. PDOS for DMSCs built using combinations of M1 transition metals (Co, Rh and Ir) with the M2 transition metals (Co, Rh and Ir) belonging to the same group in the periodic table. 122

Figure B.6. PDOS for DMSCs build using combinations of M1 transition metals (Cr, Mn and Fe) with the M2 transition metals (Rh and Ir) belonging to the same group in the periodic table. 124

Figure B.7. Distribution of $E_{(CH4*)}$ values based on their Z-score. 125

Figure B.8. Distribution of $E_{(CH3*, H*)}$ values based on their Z-score. 125

Figure B.9. Distribution of Ea values based on their Z-score. 126

List of Tables

Table 1.1. Tabulated machine learning (ML) based studies reported in the literature on transition metal (TM) based single atom (SAAs) and bimetallic alloys with best-performing (ML) model along with their accuracy in the form of root mean square error (RMSE) and mean absolute error (MAE)....................... 15

Table 1.2. Tabulated machine learning (ML) based studies reported in the literature on graphene-based single atom (SAC) and dual metal site (DMSC) catalysts with best performing ML model along with their accuracy in the form of root mean square error (RMSE), mean square error (MSE) and mean absolute error (MAE)... 18

Table 3.1. Adsorption energy of atomic oxygen (E_O) and carbon (E_C) on (111), (110), and (100). 40

Table 3.2. Optimal hyperparameter values are used in each ML model. 45

Table 4.1. Optimized hyperparameters of ML models for E_O and E_C prediction. 54

Table 6.1. Optimized hyperparameters of ML models for E_a and $E_{CH3*, H*}$ prediction. 89

Table A.1. The range of hyperparameters used in each ML model. 105

Table A.2. The formation energies of all the species with the surface used in the MKM. These are referenced to CH_4, H_2O and H_2.. 106

Table B.1. The range of hyperparameters used in each ML model. 126

List of Abbreviations

Abbreviation	**Definition**
LH	Langmuir–Hinshelwood
ER	Eley–Rideal
QM	Quantum Mechanical
NoMaD	Novel Materials Discovery
OQMD	Open Quantum Materials Database
CMR	Computational Materials Repository
DFT	Density functional theory
ML	Machine Learning
MKM	Microkinetic Model
ORR	Oxygen Reduction Reaction
HER	Hydrogen Evolution Reaction
NRR	Nitrogen Reduction Reaction
OER	Oxygen Evolution Reaction
CRR	CO2 Reduction Reaction
SAA	Single Atom Alloy
DAA	Di Atom Alloy
TAA	Tri Atom Alloy
SAC	Single Atom Catalyst
DMSC	Dual Metal Site Catalyst
LR	Linear Regression
KRR	Kernel Ridge Regression
RR	Ridge Regression
KNN	K Nearest Neighbour
SVM	Support Vector Machine
SVR	Support Vector Regression
RFR	Random Forrest Regression
ETR	Extra Tree Regression
GBR	Gradient Boost Regression
XGBR	Extreme Gradient Boost Regression
LGBM	Light Gradient Boosting Machine
GPR	Gaussian Processes Regression
NN	Neural Network
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DNN	Deep Neural Network
FNN	Feedforward Neural Network
OLS	Ordinary Least Squares
PLS	Partial Least Squares

LASSO	Least Absolute Shrinkage and Selection Operator
SISSO	Sure Independence Screening and Sparsifying Operator
RMSE	Root Mean Square Error
MSE	Mean Square Error
MAE	Mean Absolute Error
TOF	Turn-Over Frequency
DOS	Density of State
TS	Transition State
MEP	Minimum Energy Path
NEB	Nudged Elastic Band
CI-NEB	Climbing Image Nudged Elastic Band
OCM	Oxidative Coupling of Methane
NOCM	Non-Oxidative Coupling of Methane
NODH	Non-Oxidative Dehydrogenation
BEP	Brønsted-Evans-Polanyi
LDA	Local Density Approximation
GGA	Generalized Gradient Approximation
PW91	Perdew–Wang Functional
PBE	Perdew–Burke–Ernzerhof Functional
RPBE	Revised Perdew–Burke–Ernzerhof Functional
BZ	Brillouin Zone
USPP	Ultrasoft Pseudopotentials
PAW	Projector Augmented Wave
E_O	Adsorption Energy of Atomic O
E_C	Adsorption Energy of Atomic C
E_a	Activation Energy
E_{CH4*}	Initial State Adsorption Energy of CH4*
$E_{CH3*, H*}$	Final State Adsorption Energy of (CH3*, H*)