

AI Accelerated Optimization and Prediction of Key Performance Parameters in Catalytic Methane Decomposition



By

Muhammad Harram

(Registration No: 00000330672)

Department of Chemical Engineering

School of Chemical and Materials Engineering

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

AI Accelerated Optimization and Prediction of Key Performance Parameters in Catalytic Methane Decomposition



By

Muhammad Harram

(Registration No: 00000330672)

A thesis submitted to the National University of Sciences and Technology, Islamabad,

in partial fulfillment of the requirements for the degree of

Master of Science in
Chemical Engineering

Supervisor: Dr. Umair Sikandar

Co Supervisor: Dr. Muhammad Nouman Aslam Khan

School of Chemical and Materials Engineering

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)



THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS thesis written by Mr Muhammad Harram (Registration No 00000330672), of School of Chemical & Materials Engineering (SCME) has been vetted by undersigned, found complete in all respects as per NUST Statues/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: 

Name of Supervisor: Dr Umair Sikandar

Date: 26-09-24

Signature (HOD): 

Date: 26/09/24

Signature (Dean/Principal): 

Date: 26/9/24

National University of Sciences & Technology (NUST)

MASTER'S THESIS WORK

Formulation of Guidance and Examination Committee (GEC)

Name: Muhammad Harram NUST Reg No: 00000330672
 Department: Department of Chemical Engineering Specialization: Master of Science in Process System Engineering
 Credit Hour Completed: 24.0 CGPA: 3.19

Course Work Completed

S/No:	Code:	Title:	Core/Elective:	CH:	Grade:
1.	PSE-801	Process Systems Theory	Compulsory	3.0	B+
2.	PSE-852	Process Modelling and Simulation	Compulsory	3.0	C
3.	RM-898	Research Methodology	Additional	2.0	Q
4.	TEE-820	Process Intensification	Compulsory	3.0	A
5.	PSE-802	Optimization and Decision Analysis	Compulsory	3.0	B
6.	PSE-823	Advanced Process Dynamics and Control	Compulsory	3.0	B+
7.	CSE-801	Computational Fluid Dynamics	Elective	3.0	B
8.	ENE-809	Waste Water Treatment & Design	Elective	3.0	B
9.	CHE-814	Product Technology	Elective	3.0	B+

Date 25-Oct-2022

Student's Signature

Thesis Committee

1. Name: Umair Sikander (Supervisor)Department: Department of Chemical Engineering

Signature

2. Name: Muhammad Nouman Aslam Khan (Cosupervisor)Department: Department of Chemical Engineering

Signature

3. Name: Dr. Nouman Ahmad (Internal)Department: Department of Chemical Engineering

Signature

4. Name: Muhammad Ahsan (Internal)Department: Department of Chemical Engineering

Signature

Date: 25-Oct-2022

Signature of Head of Department:

APPROVAL

Date: 25-Oct-2022

Signature of Dean/Principal:



Form: TH-04

National University of Sciences & Technology (NUST)

MASTER'S THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by

Regn No & Name: 00000330672 Muhammad Harram

Title: AI Accelerated Optimization and Prediction of Key Performance Parameters in Catalytic Methane Decomposition.

Presented on: 05 Sep 2024 at: 1100 hrs in SCME (on MS Teams)

Be accepted in partial fulfillment of the requirements for the award of Master of Science degree in Process Systems Engineering.

Guidance & Examination Committee Members

Name: Dr Nouman Ahmad

Signature: [Signature]

Name: Dr Muhammad Ahsan

Signature: [Signature]

Name: Dr M. Nouman Aslam Khan (Co-Supervisor)

Signature: [Signature]

Supervisor's Name: Dr Umair Sikander

Signature: [Signature]

Dated: 25-9-24

[Signature]

Head of Department

Date 25-09-24

[Signature]

Dean/Principal

Date 25/9/24

School of Chemical & Materials Engineering (SCME)

AUTHOR'S DECLARATION

I **Muhammad Harram** hereby state that my MS thesis titled “**AI accelerated optimization and prediction of key performance parameters in catalytic methane decomposition**” is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Name of Student: Muhammad Harram

Date: September 26, 2024

PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the thesis titled “**AI accelerated optimization and prediction of key performance parameters in catalytic methane decomposition**” is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the University reserves the rights to withdraw/revoke my MS degree and that HEC and NUST, Islamabad has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature:  _____

Name: Muhammad Harram

DEDICATION

"To my very supportive, loving and caring family"

ACKNOWLEDGEMENTS

All praise and eminence are due to "ALLAH," the undisputed architect of this world, who gave us the capacity for comprehension and sparked our curiosity about the planet as a whole. Warmest welcomes to the supreme ruler of this world and the hereafter, "Prophet Mohammed (PBUH)," a source of knowledge and benefits for all of humanity as well as for Umah.

I would like to convey my profound gratitude to **Dr. Umair Sikandar and Dr Nouman Aslam Khan** for their leadership, motivation, helpful suggestions, dedication, and thorough oversight during the project. For their aid during my program, I am appreciative to them. It was a fantastic privilege and an honour to work under his direction. I shall cherish this memory for my entire life.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	IX
TABLE OF CONTENTS	X
LIST OF TABLES	XII
LIST OF FIGURES	XIII
LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS	XIV
ABSTRACT	XV
CHAPTER 1 INTRODUCTION	1
1.1 Background and Context	1
1.2 Research Objective and Question	2
1.3 Significance and Motivation	2
1.4 Scope and Limitations	2
1.5 Organization of Thesis	3
CHAPTER 2 LITERATURE REVIEW	4
CHAPTER 3 OVERVIEW ML MODELS AND OPTIMIZATION FRAMEWORK	7
3.1 Machine Learning Models	7
3.1.1 Artificial Neural Network	7
3.1.2 Ensemble Trees	8
3.1.3 Gaussian Process Regression	9
3.1.4 Regression Tree	10
3.1.5 Support Vector Machine	11
3.2 Genetic Algorithm Optimization	12
CHAPTER 4 RESULTS AND DISCUSSIONS	16
4.1 Methodology	16
4.1.1 Data Collection	16
4.1.2 Data Distribution	16
4.1.3 Performance Evaluation Criteria	18

4.1.4	Pearson Correlation	19
4.1.5	Data Pre-processing	20
4.2	GA for feature selection	20
4.3	GA based tuning of hyperparameters	22
4.4	Prediction Performance of ML models	23
4.5	Effect Of Input variables on H₂ yield (%)	26
4.6	Graphical User Interface (GUI)	31
4.7	Experimental Validation	32
CHAPTER 5 CONCLUSIONS AND FUTURE DIRECTIONS		36
5.1	Conclusions	36
5.2	Future Directions	37
REFERENCES		38

LIST OF TABLES

Table 3.1: Parameters of Genetic Algorithm	15
Table 4.1: Data Distribution of Data.....	17
Table 4.2: GA based feature selection	21
Table 4.3: GA based hyperparameters optimization.....	22
Table 4.4: Comparison of ML models	24
Table 4.5: Performance Table of Graphical User Interface for Hydrogen Yield Prediction ...	33

LIST OF FIGURES

Figure 3.1: Structure of ANN	8
Figure 3.2: Bagging in ET	9
Figure 3.3: Workflow of GPR	10
Figure 3.4: Workflow of RT	11
Figure 3.5: Schematic of SVM	12
Figure 3.6: Workflow of GA	13
Figure 4.1: Box Plots of Input and output parameters.	18
Figure 4.2: Correlation Heat map	20
Figure 4.3: A) RT training a) RT testing, B) ET training b) ET testing, C) GPR Training c) GPR testing, D) SVM training d) SVM testing.	26
Figure 4.4: 2D Partial Dependence plots of H ₂ yield (%).....	29
Figure 4.5: Partial dependence plots for Hydrogen Yield (%)	31
Figure 4.6: Graphical User Interface (GUI) for Hydrogen Yield (%) Prediction.....	32
Figure 4.7: Comparison of the influence of temperature on A) ML study and B) experimental study.....	34
Figure 4.8: Comparison of the influence of temperature on A) ML study and B) experimental study.....	35

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

CDM	Catalytic methane decomposition
ML	Machine learning
ANN	Artificial Neural Network
ET	Ensembled Tree
GPR	Gaussian Process Regression
RT	Regression Tree
SVM	Support Vector Machine
R^2	Coefficient of correlation
RMSE	Root Mean Square Error

ABSTRACT

The depletion of conventional fuels and the urgent need for clean and sustainable energy sources have driven research into H₂ production through catalytic methane decomposition. This study aimed to use novel machine learning (ML) approaches to enhance and predict H₂ yield using Artificial Neural Network (ANN), Ensembled Tree (ET), Gaussian Process Regression (GPR), Regression Tree (RT), and Support Vector Machine (SVM). A two-step approach involving feature selection and hyperparameter optimization was employed to enhance the models' performance for H₂ yield. The coefficient of correlation (R²) and Root Mean Square Error (RMSE) were used to evaluate model performance. ET performed excellent with R² of 0.929 (training) and 0.933 (testing) however GPR exhibited exceptional performance, achieving a perfect training R² of 1.00 and low RMSE 0.00026. Furthermore, partial dependence plots (PDPs) were utilized to assess the impact of catalyst properties and reaction conditions on H₂ yield. Temperature impacts H₂ directly, while time shows an inverse relationship. Various catalysts and catalysts structure exhibited distinct behaviors, and the Average surface area demonstrated a direct linear relationship with H₂ yield. These findings contribute to the understanding of catalytic methane decomposition and provide insights for optimizing H₂ production processes using ML models. Given the depletion of conventional fuels, H₂ has emerged as a crucial alternative energy source due to its clean and sustainable nature. The ability to accurately predict H₂ yield using ML models opens new avenues for advancing H₂ production technologies and meeting the growing global energy demand while mitigating environmental concerns.

Keywords: Catalytic methane decomposition, Machine learning, H₂ yield

CHAPTER 01: INTRODUCTION

1.1 Background and Context

The process of generating hydrogen from catalytic methane decomposition (CDM) is a significant area of research because of its potential to provide a sustainable and clean energy source. This method involves the decomposition of methane (CH₄) into carbon (C) and hydrogen (H₂) using a catalyst, which is a key step in the decarbonization of fossil fuels and the production of hydrogen for various applications, including fuel cells and synthetic fuels. The catalytic decomposition of methane is a thermochemical reaction that can be represented by the following equation:



This process releases a significant amount of energy, making it an attractive method for hydrogen production. The reaction's enthalpy change (ΔH°) is approximately +74.8 kJ/mol, indicating the exothermic nature of the reaction. Various catalysts have been explored for CDM, including iron-based catalysts, which have shown promising results in terms of efficiency and selectivity. For instance, iron-containing catalysts have been used for methane decomposition, demonstrating the accumulation of filamentous carbon, which can be beneficial for the production of nanostructured carbon materials. In addition, catalysts such as Ni/SiO₂ and Fe/SiO₂ have been studied for their role in producing hydrogen and filamentous carbon via methane decomposition. The development of high-temperature catalysts, such as carbon capacious Ni-Cu-Al₂O₃ catalysts, has also been investigated for their performance in methane decomposition. These catalysts have shown potential for high-temperature applications, which is crucial for the efficient production of hydrogen. The use of molten salt-promoted Ni-Fe/Al₂O₃ catalysts has been explored for methane decomposition, highlighting the potential of such catalysts in enhancing the efficiency of process. Furthermore, ceria-modified iron catalysts have been studied for their role in methane decomposition, demonstrating the effectiveness of such catalysts in producing hydrogen and nanostructured carbon. In summary, the catalytic decomposition of methane is a promising method for hydrogen production, and various catalysts are being explored to optimize the process. The development of efficient and selective catalysts is crucial for the

commercialization of this technology, which could play a significant role in the transition toward cleaner and more sustainable energy sources.

1.2 Research Objective and Question

Objective: The primary objective of this research is to develop a machine learning model capable of accurately predicting hydrogen yield in catalytic methane decomposition (CDM) processes.

Research Questions:

- Can machine learning techniques effectively capture the complex relationships among catalyst properties, process conditions, and hydrogen yield in CDM?
- What are the most influential factors affecting hydrogen yield in catalytic methane decomposition?
- How can machine learning models be optimized to provide reliable predictions of hydrogen yield in CDM under various operating conditions?
- How do different types of catalysts and process parameters affect the performance of machine learning models for predicting hydrogen yield in CDM?

1.3 Significance and Motivation

The significance of this research lies in its potential to enhance the efficiency and sustainability of hydrogen production through catalytic methane decomposition. By employing machine learning techniques, we provide a predictive tool that can optimize CDM processes, leading to increased hydrogen yield and reduced energy consumption. Such advancements are crucial for the widespread adoption of hydrogen as a clean energy source, contributing to global efforts for decarbonization and environmental sustainability.

1.4 Scope and Limitations

Scope:

- This research will focus on developing machine learning models specifically tailored for predicting hydrogen yield in catalytic methane decomposition.
- Various catalyst types and process parameters will be considered to assess their impact on the model performance.

- This study will involve both experimental data and computational simulations to train and validate the machine learning models.

Limitations:

- The accuracy of machine learning predictions may be determined by the availability and quality of data and the complexity of the underlying chemical reactions.
- The models developed in this study may have limited generalizability outside the specific conditions and catalysts considered.
- External factors such as reactor design and catalyst degradation over time may not be fully accounted for in the predictive models.

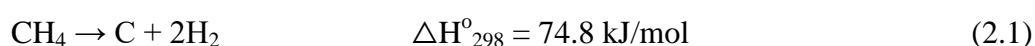
1.5 Organization of Thesis

The remainder of this thesis is organized as follows:

- Chapter 2 provides an in-depth review of the relevant literature on catalytic methane decomposition (CDM)
- Chapter 3 outlines the theoretical framework, encompassing the fundamental principles of CDM, machine learning techniques for predicting hydrogen yield, and optimization methods for enhancing CDM efficiency.
- Chapter 4 presents the results and discussions, detailing the effect of parameters in CDM, such as catalyst properties and process conditions. This chapter also covers data collection, model development using ML algorithms, integration of optimization techniques, and validation procedures for predicting hydrogen yield in CDM.

CHAPTER 02: LITERATURE REVIEW

H₂ is considered the energy of the future due to its clean burning properties and high calorific value. H₂ is the lightest element and approximately accounts for 74% of the composition of the universe, making it the most abundant element. H₂ production methods include water dissociation, fossil fuel conversion, and biological means. H₂ can be produced directly via Water dissociation, but this method is energy intensive. Industrially, gasification and reforming are commonly used to produce a gas called syngas, which is a mixture of H₂ and carbon monoxide and can be utilized to produce valuable chemical products. Natural gas steam reforming is a preferred method but poses an environmental concern because of its release of large amounts of carbon dioxide. Conventional fuel-based H₂ production results in the emission of large amounts of CO₂ [1]. Commercial H₂ production currently relies on processes such as Steam Reforming of Methane. However, Carbon Dioxide (CO₂) emissions from Steam Reforming of Methane, even with CO₂ capture, are found to be higher compared to Catalytic Methane Decomposition. Catalytic Methane Decomposition has lower CO₂ emissions and produces valuable by-product carbon, making it economically competitive with SRM. Catalytic Methane Decomposition has an advantage over electrolysis water splitting as it can utilize already present natural gas services unlike electrolysis which relies on the availability of renewable energy and cost. Limited availability of renewable electricity restricts the widespread use of electrolysis [2].



The reaction of catalytic methane decomposition above is considered an attractive alternative for producing clean H₂ and valuable carbon material. It is favored for its simple process flow and high efficiency. A significant advantage is that this reaction does not generate carbon dioxide, making it environmentally appealing, particularly considering global warming concerns caused by excessive carbon dioxide emissions. The resulting products, solid carbon, and usable H₂ can be easily separated. The obtained solid carbon can be utilized as a raw material in various applications, including plastic production and the metallurgical industry [3, 4]. The reaction is preferred to be carried out at temperatures higher than 1073 K because it facilitates high conversion of methane which is further used for various applications. In some cases, an H₂ stream with nearly hundred percent concentration is required, necessitating even higher temperatures, around 1273 K. However, one of the key problems of the catalytic

methane decomposition process at high temperatures is catalyst deactivation induced by the accumulation of carbon and sintering. This is because the reaction requires high temperatures greater than 873 K. To address these issues, researchers created and tested a variety of supported metal-based catalysts, with the goal of overcoming catalytic deactivation and improving overall process performance [5].

One significant challenge in catalytic methane decomposition is understanding the reaction kinetics, specifically the rate of hydrogen (H_2) and Carbon (C) production. This challenge arises from the diverse chemical compositions of different catalysts, which can affect the reaction kinetics. Understanding and characterizing these kinetics is crucial for optimizing the process and selecting the most efficient catalysts [6]. To address the challenge of understanding the kinetics of catalytic methane decomposition, an artificial intelligence modeling approach can be employed. This method allows for a better understanding of the process parameters. Irregular and complex interactions between input and output parameters are common in real-world systems. By gaining a deeper understanding of these non-linear relationships, it becomes possible to optimize process operations, establish theoretical frameworks, automate processes, and facilitate upscaling. The utilization of artificial intelligence modeling provides a valuable tool for overcoming this challenge and enhancing the efficiency of catalytic methane decomposition [7]. Artificial intelligence modeling has been extensively used in various catalytic processes. Examples include Sulfur removal [7], steam reforming of methanol, steam reforming of glycerol [8, 9], Biomass air gasification [10], shift reaction of water and gas [11], and waste palm oil steam gasification [12]. In their study, Nasr et al. [13] utilized artificial neural networks (ANNs) in a back-propagation configuration to develop a predictive model for bio hydrogen production. They found a significant correlation between the experimental and predicted bio hydrogen production results. This suggests that the ANN approach effectively predicts and models bio hydrogen production in their study. Ayodele et al. [14] utilized ANNs have been used to forecast the conversion of CH_4 , CO_2 , and the syngas ratio in methane dry reforming reactions using Co catalysts supported by Sm_2O_3 and CeO_2 . This application of ANN allows for accurate predictions of these key parameters, aiding in the optimization and understanding of the methane dry reforming process using these specific catalysts.

This study evaluates the performance of different machine learning (ML) models namely Artificial Neural Network (ANN), Ensemble Tree (ET), Gaussian Process Regression (GPR),

Regression Tree (RT), and Support Vector Machine (SVM) in predicting H₂ yield (%) from catalytic methane decomposition. It introduces Genetic Algorithms (GAs) for model optimization, leveraging natural evolution principles for improved ML model performance. By combining ML models, GA based hyper parameters tuning and feature selection, the study aims to enhance H₂ yield prediction accuracy. The analysis highlights the superior predictive capabilities of certain ML models, with ET exhibiting excellent performance and GPR showing exceptional results, achieving a perfect training R² of 1.00 and low RMSE 0.00026. Additionally, the study employs partial dependence plots (PDPs) to understand the impact of catalyst properties and reaction conditions on H₂ yield. Key findings include the direct influence of temperature on H₂ yield and an inverse relationship with time. Moreover, different catalysts and catalyst structures display distinct behaviors, while the average surface area demonstrates a direct linear relationship with H₂ yield.

CHAPTER 03: OVERVIEW ML MODELS AND OPTIMIZATION FRAMEWORK

3.1 Machine Learning Models

This section provides a broad review of all the ML models developed for this thesis.

3.1.1 *Artificial Neural Network*

An Artificial Neural Network (ANN) is a machine learning model which takes its inspiration from the working of the human brain. It consists of interconnected artificial neurons organized in layers. ANNs are used for tasks like classification and regression. Neurons receive inputs, apply transformations, and pass outputs to the next layer. Connections between neurons have weights determining signal strength. ANNs have input, hidden, and output layers. Hidden layers extract features from data. During training, an ANN adjusts connection weights using optimization algorithms like backpropagation. It minimizes the difference between predicted and desired outputs over iterations. Trained ANNs make predictions or classifications on new data. Input passes through layers, and the network applies learned weights to produce an output. ANN success depends on learning from training data to generalize for accurate predictions on unseen data [15]. The application of ANNs extends to predicting outcomes based on input parameters, specifically in H₂ yield prediction. During this process, input data related to the catalytic process traverses through the layers of the ANN. The hidden layers, skilled in feature extraction, identify patterns and relationships among the input parameters temperature, time, catalyst, CS, Nickel (%), BET, and structural features. Applying the learned weights, the network generates an output representing the predicted H₂ yield (%). Figure 1 shows the schematic of ANN

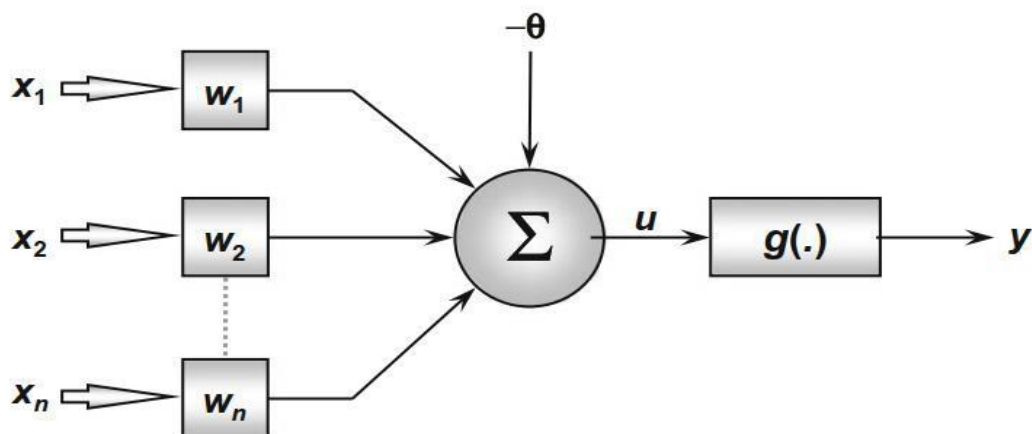


Figure 3.1: Structure of ANN

3.1.2 Ensemble Trees

Ensemble Trees (ET), or decision tree ensembles, enhance machine learning predictions by combining multiple trees. Each tree is trained on a subset of data and random features, promoting diverse learning. During prediction, the ensemble aggregates individual tree outputs, using majority voting for classification and averaging for regression. ET operates on the "wisdom of the crowd" principle, compensating for individual weaknesses and biases, resulting in a more accurate and robust model. Valuable in classification, regression, and anomaly detection, ET's ability to capture complex relationships makes it a widely employed tool, significantly improving predictive accuracy in various machine learning applications [16]. ET discerns intricate patterns and relationships within the input parameters including temperature, time, catalyst, CS, Nickel (%), and BET. Each decision tree, trained on a subset of data and random features, contributes to the collective prediction. By combining diverse insights from individual trees, the ensemble compensates for biases and weaknesses, producing a comprehensive output that represents the predicted H₂ yield (%) with improved accuracy.

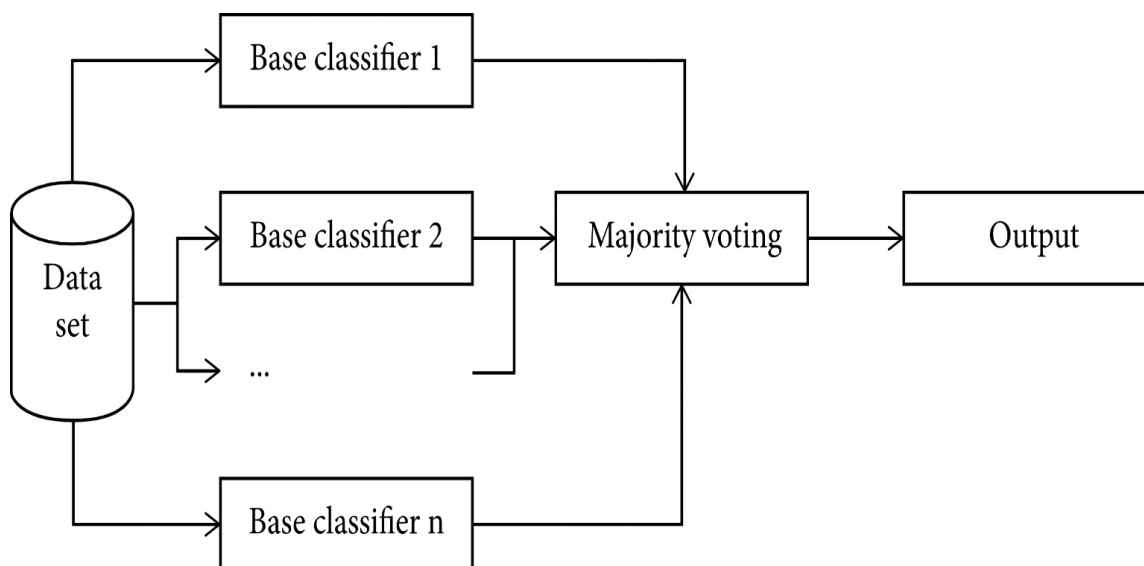


Figure 3.2: Bagging in ET

3.1.3 Gaussian Process Regression

Gaussian process regression (GPR) is a statistical machine learning technique that can be used to describe and predict continuous variables. It represents the association between input and output variables as a distribution function, without assuming a specific underlying distribution. It defines a mean function and a covariance function to represent the expected value and similarity between input points, respectively. During training, the mean and covariance function parameters are estimated from observed data using maximum likelihood. The estimated parameters allow making predictions for new input points by calculating the conditional distribution of the target variable. This distribution provides a predicted mean value and uncertainty estimate. GPR is flexible, accommodating complex relationships in data and providing uncertainty estimates. However, it can be computationally expensive for large datasets due to inverting covariance matrices. Approximation techniques can be used to address scalability issues [15]. GPR models the relationship between input parameters (temperature, time, catalyst, CS, Nickel %, BET, structural features) and H₂ yield (%), providing predictions along with uncertainty estimates. GPR's probabilistic nature allows it to adapt to data variations, offering valuable insights into both predictions and their associated confidence levels.

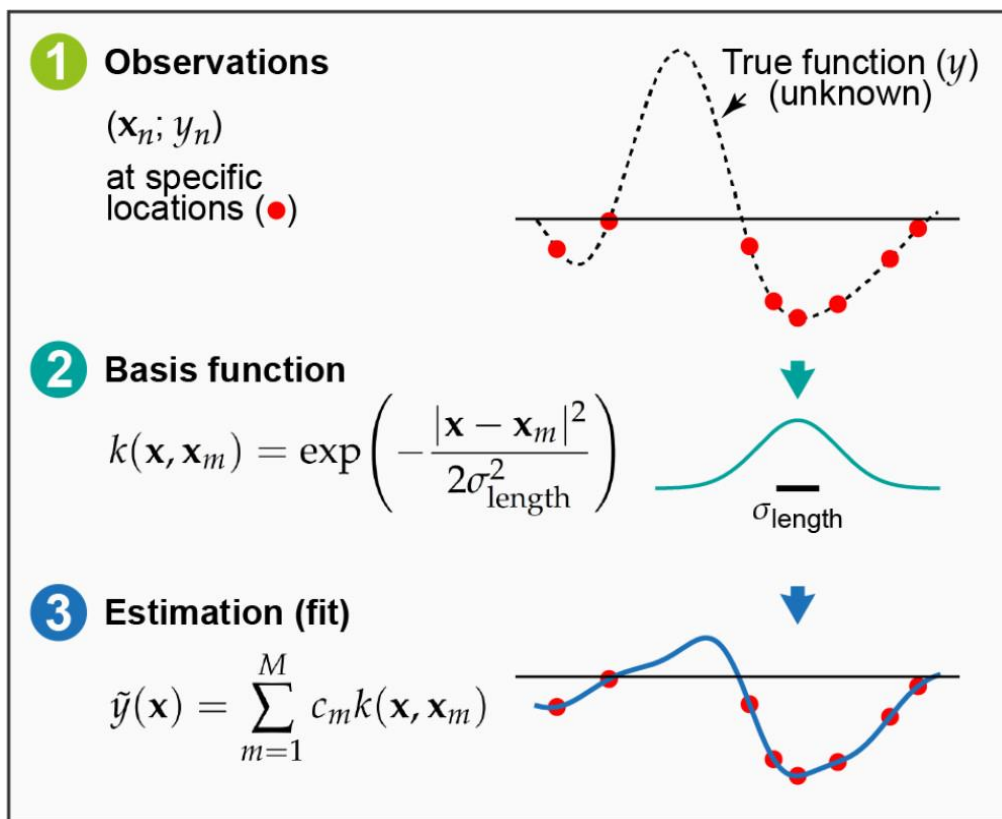


Figure 3.3: Workflow of GPR

3.1.4 Regression Tree

A regression tree (RT) is a machine learning algorithm used to predict continuous numerical values. It divides the input space into regions and assigns a constant value to each region as the prediction. The tree is constructed by repeatedly dividing the dataset based on the input features which allows the minimization of variance of objective variable within each subset. The tree starts from the root node which comprises of the entire dataset, the algorithm then selects the best feature and split point to create branches and nodes. This process is repeated until a condition, such as attaining a maximum tree depth or a minimum number of samples per leaf, or a stopping criterion is reached. To make predictions, new instances traverse the tree based on their feature values, with the predicted value being the constant associated with the leaf node. RT can capture complex relationships but are prone to overfitting. Techniques like pruning, regularization, or ensemble methods can address overfitting [17]. RT for H₂ prediction will use input parameters (temperature, time, catalyst, CS, Nickel %, BET, and structure) to recursively split data into subsets, capturing complex dependencies for accurate H₂ yield (%) forecasts.

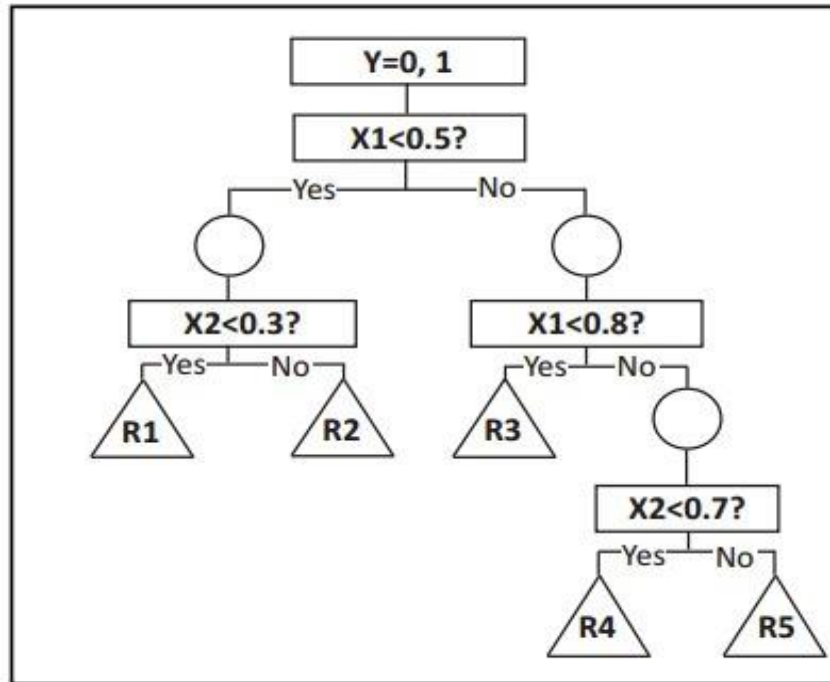


Figure 3.4: Workflow of RT

3.1.5 Support Vector Machine

Support Vector Machines (SVMs) are strong supervised machine learning algorithms which can be used for both classification and regression problems. SVM seeks the best hyperplane or decision boundary for separating data points of distinct classes or approximating a regression function. SVMs accomplish this by utilizing kernel functions to map the data into a higher-dimensional feature space. The procedure maximizes the margin between the hyperplane and the nearest data points of distinct classes, with the support vectors determining the decision boundary. SVMs are capable of handling non-linearly separable data and are resistant to overfitting. They necessitate careful hyperparameter adjustment, but they provide flexibility and effectiveness in a variety of applications [18]. SVM for H₂ prediction utilize input parameters (temperature, time, catalyst, CS, Nickel %, BET) to find a hyperplane that maximally separates data points, optimizing the margin between classes and facilitating accurate H₂ yield (%) predictions based on the input features.

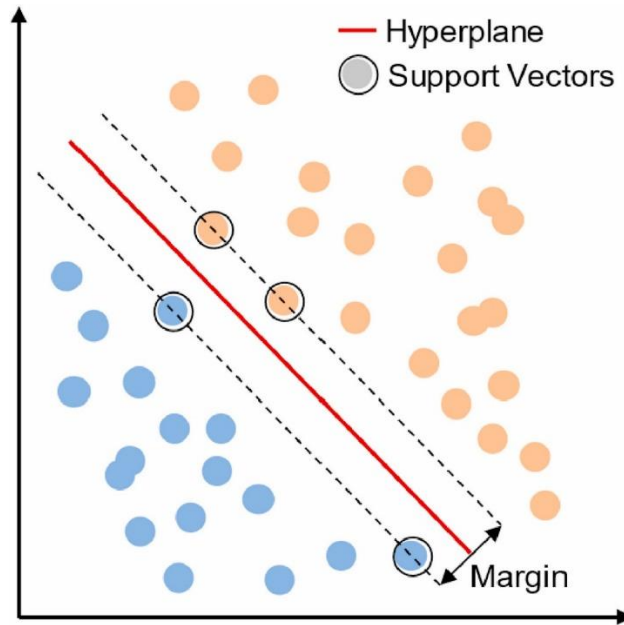


Figure 3.5: Schematic of SVM

3.2 Genetic Algorithm Optimization

Genetic Algorithm (GA) draws inspiration from nature's selection process to address problem-solving effectively. Its objective is to assist a population in adapting to dynamic environmental changes. Each individual within the population possesses distinct characteristics, necessitating the refinement of beneficial traits while eliminating redundant and detrimental qualities. GA operates as an evolutionary algorithm, incorporating evolutionary programming and strategies. The initiation of the process involves the creation of a population through random generation. [19]. The operators are crucial in the GA, which are explained below.

Numerous studies have employed Genetic Algorithms (GAs) to assess solutions for intricate problems characterized by inconsistent, non-uniform decision variables, and other complexities. In the course of searches, GA organizes potential solutions into data structures resembling chromosomes, forming a community of these genetic entities. A crucial aspect involves utilizing a fitness function that assigns a scalar reward or incentive to each solution. With a well-designed plan and evaluation function, GA can anticipate honest responses. The process begins by generating an initial solution consisting of a predetermined set of threads or chromosomes, representing the population size. Subsequently, each option in the initial population undergoes analysis using a payout function, wherein superior responses receive higher rewards, while the remaining options receive comparatively lesser rewards. The

subsequent generation is formed using GA operators, including crossover and mutation. This iterative process continues until an optimal solution (or solutions) is identified, the specified number of iterations or populations is reached, or the discrepancy among solutions falls below a predefined limit. Figure 6 illustrates the schematics of the GA process. [20-22]. The aspects of GA are described briefly below:

1. Representation in a genetic algorithm involves depicting individuals in a population through chromosomes. The chosen representation scheme dictates the problem structure and determines the applicable genetic operators. Gene sequences are formed using specific alphabets, such as binary numbers (0 and 1) or value numbers. Research indicates that chromosomes encoded with real values result in more efficient GAs and improved solutions.[23].

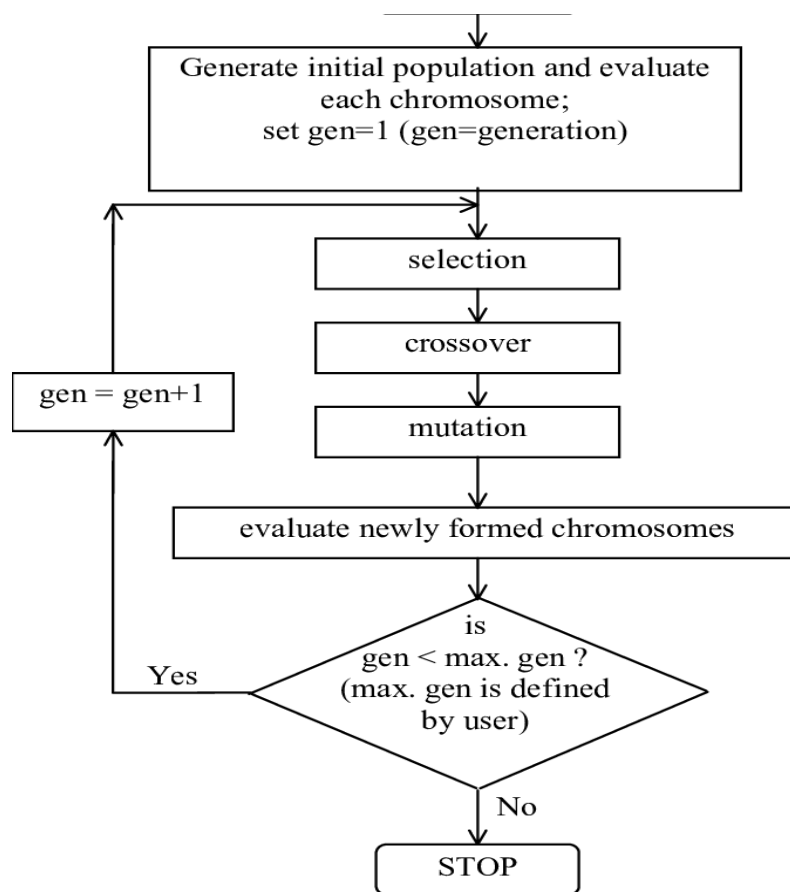


Figure 3.6: Workflow of GA

2. In GA, generations are formed by selecting individuals from the previous generation, where selection is based on each entity's likelihood of being chosen for reproduction in the next generation, determined by their fitness value. Selection methods include roulette wheel selection, scale techniques, competitions, elitist systems, and rating scales. Common to these approaches is assigning individuals a chance of selection, accomplished through a roulette wheel, logarithmic ranking, or geometrical priority.

3. Genetic operators play a crucial role in manipulating populations across generations to achieve optimal outcomes. Preserving adaptive traits from prior generations is essential. The selection operator focuses on choosing parents with high fitness values for offspring reproduction, ensuring that chromosomes with superior fitness contribute to the next generation. Another key operator is crossover, which combines parent features during reproduction, creating offspring with traits inherited from both parents. The mutation operator prevents critical data loss by randomly altering one or more individuals in a group, promoting genetic diversity. Additionally, a binary coded gene is generated by reversing the gene sequence from 0 to 1.

4. To initiate the GA process for finding the optimum solution, an initial population is required. This population can be created by generating random responses within the specified upper and lower bounds of variables. Alternatively, the starting population can be seeded with established best approaches, enhancing the overall quality. The remaining population can then be randomly generated.

5. Once a specific termination met the criteria of termination than the GA operations are terminated. The requirement of termination can be any or a mix of the following: (1) The number of generations approaches a predefined maximum value. (2) The population adapts on a unique solution. (3) The difference between solutions becomes less than a specified criterion. (4) The ideal solution does not grow over a certain no. of generations. (5) The evaluation values meet an acceptable level [22-24].

- Commencing with the creation of a randomized population set.
- Evaluating features through the utilization of a fitness function.
- In instances where the objective function's criteria were unmet, a new population batch was generated by refreshing the Genetic Algorithm (GA). GA operators were instrumental in this population generation process, as outlined in Table 1, specifying the parameters used for optimizing ML models in this study.
- This iterative process persists until the attainment of the most optimal fitting function.

Table 3.1: Parameters of Genetic Algorithm

GA Parameters	Generation	Crossover Probability	Crossover	Elite Count	Size Population	Type Population	Mutation	Probability of Mutation
Values	100	0.8	Scattered	3.95	50	Bitstring	Uniform	0.1

CHAPTER 04: RESULTS AND DISCUSSIONS

4.1 Methodology

4.1.1 Data Collection

In this section, experimental data were extracted from relevant published papers, including the catalytic methane decomposition into H₂. A total of 943 data points from 64 relevant published papers were collected comprehensively on features that significantly influence H₂ yield (%). Major features, namely temperature, time (min), catalyst, catalyst structure (CS), and average surface area (BET), were considered as input parameters. H₂ yield (%) was designated as the objective function of the catalytic methane decomposition process.

4.1.2 Data Distribution

Boxplots are used to describe the distribution of data. Boxplots show the mean, median, quartile, minimum, and maximum values of the data. Figure 4.1 exhibits the box plots and violin plots of the input and output parameters. In Figure 4.1 the rectangular box shows where the 50% of the data lies, the horizontal line present inside the box represents the median value of the data, the vertical line extended from the rectangular box shows the range of the data except the outliers, the outliers are represented as separate data points beyond the vertical line. Violin plots encompassing the box plots display the distribution, central tendency, and variability of data of input and output parameters. They provide a comprehensive visualization of continuous data, facilitating comparisons across parameters. For the data set of catalytic methane decomposition, temperature ranges from 500°C to 850°C, time ranges from 0 to 1130 min, nickel data ranges from 0 to 77%, Catalyst and CS are categorical variables where Catalysts type ranges from 1 to 27, and structure ranges from 1 to 8. Each numerical value within these ranges represents a distinct category for the respective variable. The specific names corresponding to each category are provided in supplementary files. BET ranges from 10.6 to 738.8, and output parameter H₂ yield (%) ranges from 0 to 96% with mean of 42.9138 further distribution of data can be found in table 4.1.

Table 4.1: Data Distribution of Data

	Temp (°C)	Time (min)	Nickel (%)	Average Surface Area (m²/g)	H₂%
Mean	650.901379	193.612937	35.662778	161.748144	42.913874
Std	74.285227	168.357356	19.608449	134.970620	21.572582
Min	500.000000	0.000000	0.000000	10.600000	0.000000
25%	600.000000	68.500000	20.000000	42.100000	27.694268
50%	650.000000	153.000000	40.000000	147.700000	40.000000
75%	700.000000	281.500000	50.000000	225.000000	59.000000
Max	850.000000	1130.000000	77.000000	738.000000	96.000000

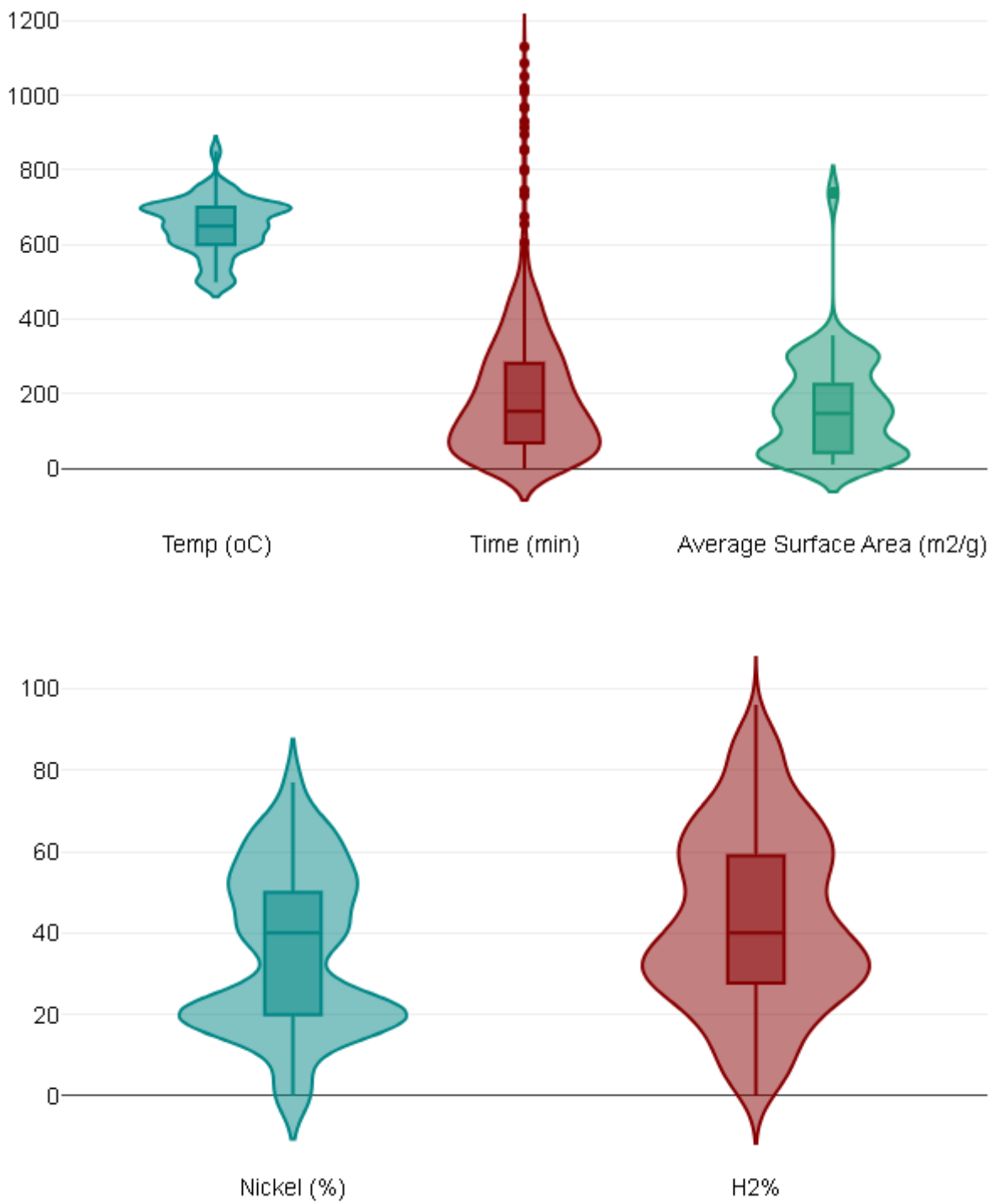


Figure 4.1: Box Plots of Input and output parameters.

4.1.3 Performance Evaluation Criteria

To check the accuracy and performance of ML models some statistical parameters are used. In this study coefficient of correlation (R^2) and Root Mean Square Error (RMSE) were

used to evaluate the accuracy of ML models. The equation for both parameters is given below.

$$R^2 = \frac{[\sum_{i=1}^n (y_{p,i} - \bar{y}_p) \times (y_{a,i} - \bar{y}_a)]^2}{\sum_{i=1}^n (y_{p,i} - \bar{y}_p)^2 \times \sum_{i=1}^n (y_{a,i} - \bar{y}_a)^2} \quad (4.1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{p,i} - y_{a,i})^2} \quad (4.2)$$

Where, \bar{y}_a = observed output mean, $y_{a,i}$ =observed output, \bar{y}_p = predicted output mean, $y_{p,i}$ =Predicted output, n =total data points.

4.1.4 Pearson Correlation

Pearson correlation quantifies the relationship of input variables to output variables. In this section correlation of Temperature, time, catalyst, CS and BET was analyzed with H₂ yield (%). Input parameters with values closer to 1 have a strong relationship with H₂ yield (%) and parameters with correlation value closer to 0 have weak relationship with H₂ yield (%). The positive and negative signs merely describe the direction of correlation of input variables with H₂ yield (%). According to figure 4.2, structure and BET show positive correlation to H₂ yield (%) with contribution of 8.9% and 27% respectively. Temp (°C), Time (min), Nickel, Catalysts contribute 1.4%, 37%, 8.9% and 13% to H₂ yield (%) but with negative association.

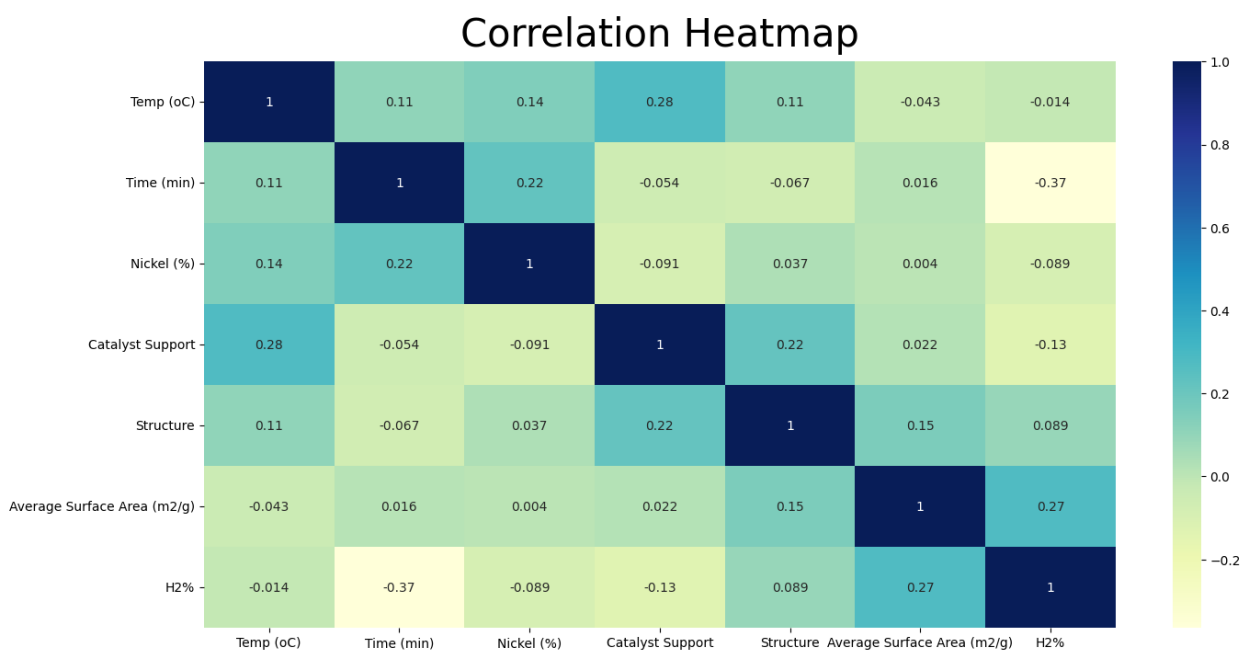


Figure 4.2: Correlation heat map

4.1.5 Data Pre-processing

After analyzing the data distribution, data preprocessing techniques were applied to prepare the data for machine learning models. Several steps were taken to ensure data quality and enhance model performance. Firstly, missing values were handled by either imputing them with appropriate measures like mean or median, or by removing the corresponding instances. This ensured that the data used for training was complete and reliable. Next, outliers were identified and treated to prevent them from disproportionately influencing the model. Feature scaling techniques like standardization or normalization were employed to bring features to a similar scale, avoiding biases in model training. Feature selection was employed via Genetic Algorithm to reduce the number of features and eliminate redundant information. These preprocessing steps were crucial in preparing the data, improving its quality, and ensuring that the machine learning models can effectively learn from it.

4.2 GA for feature selection

Feature selection is a crucial step in ML model development where input features are selected to create an effective model. It helps avoid overfitting and excessive computational complexity by including only relevant features. However, selecting features based solely on importance can be insufficient. Feature selection improves data quality by reducing

dimensionality and algorithm speed, eliminating unnecessary and redundant features [25]. Forward and backward selection are two typical ways to feature selection. Forward selection gradually introduces factors that reduce error, whereas backward selection begins with all variables and eliminates those that have no effect on error. The filter, wrapper, and hybrid approaches are three commonly used feature selection techniques. Preprocessing processes are used in filter algorithms to pick features based on correlation with the output. Wrapper approaches analyze subsets of features using a specified classification algorithm. Wrapper approaches yield better results than filter methods in general, but they demand more computer resources [26].

Feature selection using GAs aims to find the most relevant features that significantly contribute to a machine learning model's predictive performance. GAs efficiently explores a large search space by emulating the process of natural selection. Different feature subsets are represented as a population, and their evaluation is based on a fitness criterion. This criterion can be classification accuracy, regression error, or a specific performance metric. The fitness function quantifies the model's performance with the selected features. GAs evolves the population by selecting the fittest feature subsets for reproduction and applying genetic operators like crossover and mutation. This process explores various feature combinations and converges towards subsets that yield optimal performance [27]. Table 4.2 shows the GA based selected features of all the ML models. SVM and ANN show maximum performance when all the features are selected. GA based ET, GPR, RT have discarded features which increase the complexity and accuracy of the model.

Table 4.2: GA based feature selection

Model Type	Features Selected for GA
ET	Temp, time, Ni, CS, BET
GPR	Temp, time, Ni, catalyst, CS, BET

RT	Temp, time, Ni, BET
SVM	Temp, time, Ni, CS, BET
ANN	Temp, time, Ni, CS, BET

4.3 GA based tuning of hyperparameters

Hyperparameter tuning is a vital step in machine learning to optimize model performance. Hyperparameters, set by the user before training, impact various aspects of the learning process and greatly affect model outcomes. The key reasons for hyperparameter tuning are Model performance, generalization, Model complexity, speed, efficiency and model interpretability [28]. By using GA for hyperparameter tuning, it becomes possible to explore a wide range of hyperparameter combinations efficiently. GA can handle both continuous and discrete hyperparameters, as well as interactions between them. They offer the advantage of global search and can discover non-intuitive combinations that may improve model performance. The hyperparameters optimized via GA for ET are number of cycles, learning rate, and methods, for GPR optimized parameters were Basic function, Kernel function and sigma, for RT optimized parameters were minimum leaf size and surrogate, for SVM optimized parameters were Box Constraint, Kernel Scale, Epsilon and Kernel Function. Table 4.3 shows all the names, ranges and optimized values of the tuned hyperparameters for ML models via GA.

Table 4.3: GA based hyperparameters optimization

ML Methods	Parameters	Parameters (Range)	Optimized Values for GA
	Number of learning cycles	[10-500]	141.7757

ET	Learning rate	[0.001-1]	0.1044
	Methods	[LSBoost, Bag]	Bag
GPR	Basic Function	[Constant, Zero, Linear]	None
	Kernel Function	[Squared exponential, Ardexponential]	Ardexponential
	Sigma	[0.001-219.517]	134.9696
RT	Minimum Leaf Size	[1-377]	1
	Surrogate	[on, off]	on
SVM	Box Constraint	[0.001-1000]	965.1073
	Kernel Scale	[0.001-1000]	0.8642
	Epsilon	[0.002463-2446.25]	1.8058
	Kernel Function	[Polynomial, gaussian, linear, cubic]	gaussian

4.4 Prediction Performance of ML models

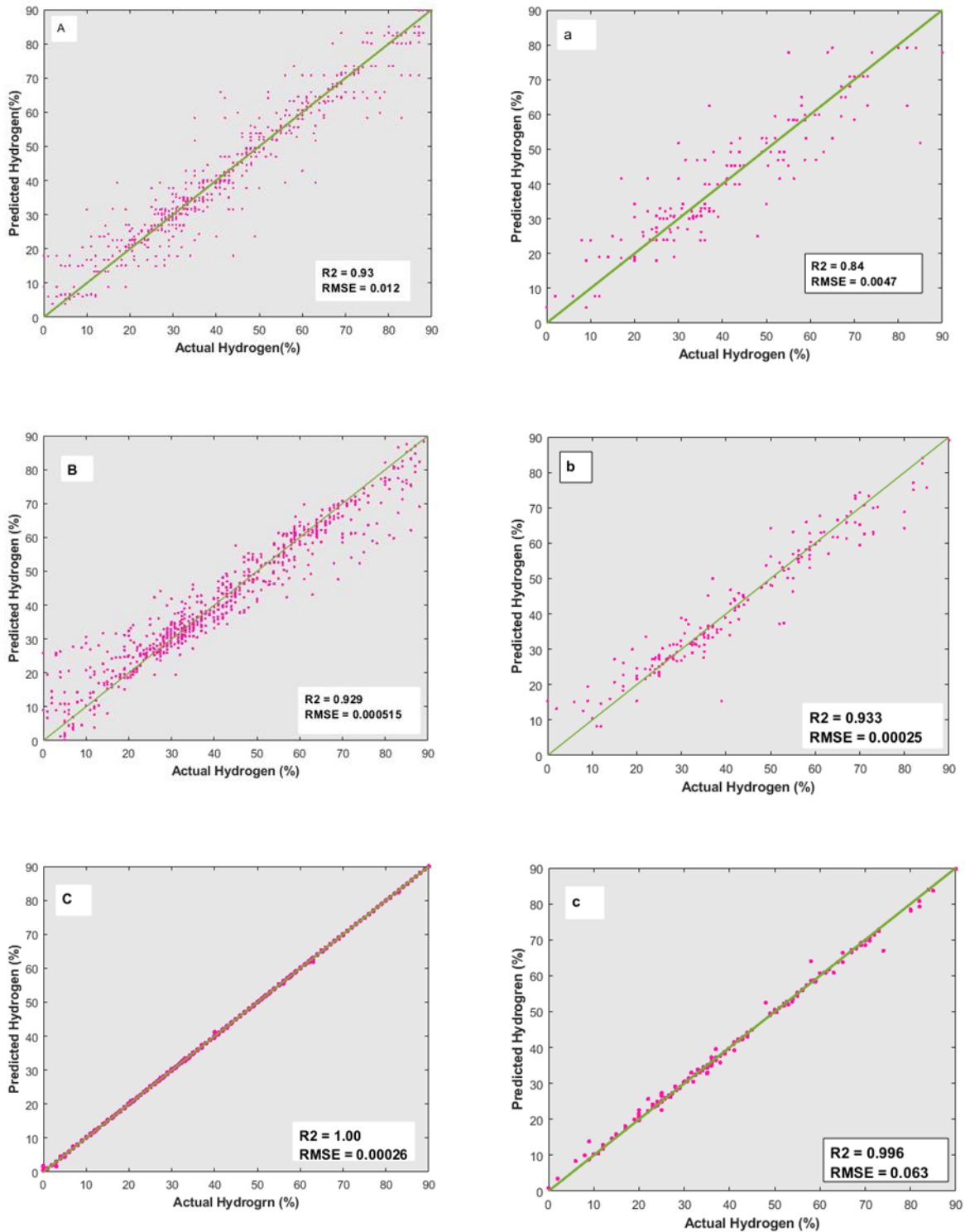
The prediction of H₂ yield (%) from catalytic methane decomposition was carried out using various machine learning models, including ANN, ET, GPR, RT, and SVM. To improve the performance and accuracy of these models, a two-step approach was followed. GA was employed for feature selection, discarding irrelevant features and optimizing the values of hyperparameters. This ensured that the models focused on the most informative

input features and had the best parameter settings. After the feature selection and hyperparameter tuning, the ML models were trained using 80% of the available data, and the remaining 20% was used for testing. The performance of the models was evaluated using the R^2 and RMSE. Table 4.4 demonstrates the satisfactory performance of all models. ET model achieved a training and testing R^2 of 0.929 and 0.933, respectively, with RMSE values of 0.000515 and 0.00025. GPR exhibited excellent performance with training R^2 of 1.00 and RMSE of 0.00026, and testing R^2 of 0.996 and RMSE of 0.063. RT showed satisfactory results with training R^2 of 0.93 and RMSE of 0.012, and testing R^2 of 0.84 and RMSE of 0.0047. SVM model showed a strong linear relationship between predicted and actual H_2 yield (%), with high training and testing R^2 of 0.993 and 0.992 respectively. However, the SVM model also had relatively high RMSE values of 12.37 and 9.287 for training and testing respectively, indicating significant deviations between predicted and actual values. This discrepancy between high R^2 and high RMSE in SVM can be attributed to the nature of the problem and the distribution of the data. ANN attained an R^2 of 0.95. Interestingly, ANN and SVM demonstrated better results when trained with all input features, while the other models discarded some features to reduce complexity and enhance performance. This can be attributed to the complexity and non-linear relationships present in the data. By including all features, ANN and SVM were able to capture the intricate relationships between the input variables and the H_2 yield (%), leading to improved performance.

Table 4.4: Comparison of ML models

Model	Train R^2 (GA)	Train RMSE (GA)	Test R^2 (GA)	Test RMSE (GA)
ET	0.929	0.000515	0.933	0.00025
GPR	1.00	0.00026	0.996	0.063
RT	0.93	0.012	0.84	0.0047
SVM	0.993	12.37	0.992	9.287

Figure 4.3 shows the difference between actual H₂ yield (%) and predicted H₂ yield (%). The actual data is depicted by a solid line in the figures, while the predicted values from the proposed machine learning models are represented by red dotted circles surrounding the solid line.



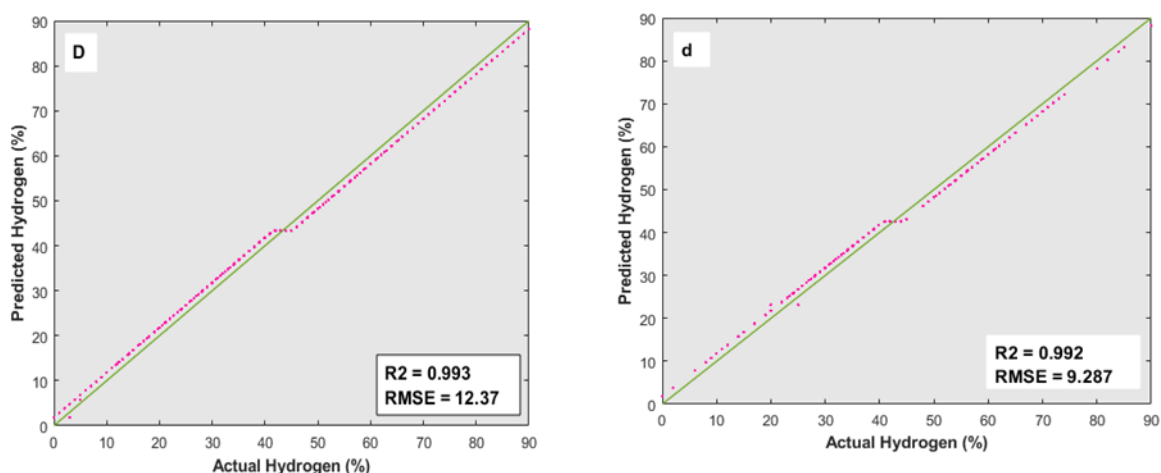


Figure 4.3: A) RT training a) RT testing, B) ET training b) ET testing, C) GPR Training c) GPR testing, D) SVM training d) SVM testing.

4.5 Effect Of Input variables on H₂ yield (%)

The effect of input variables on output variable H₂ yield (%) is determined by partial dependence plots (PDPs). PDPs are used to interpret the underlying relationship and importance of parameters on the output. The effect of selected features of the best performing model (GPR) are determined. In thermo-catalytic decomposition of methane, the temperature plays a crucial role in influencing the H₂ yield (%). Figure 4.4A shows a characteristic trend as the temperature is elevated, the H₂ yield (%) increases, reaching its peak around 650°C. This phenomenon can be attributed to the enhanced kinetics of methane decomposition at higher temperatures, promoting the dissociation of methane into H₂ and carbonaceous species. The decline in H₂ yield (%) after 650°C can be attributed to the sintering of active sites, and elevation after 750°C is owed to the shift of thermos-catalytic methane decomposition to thermal methane decomposition.

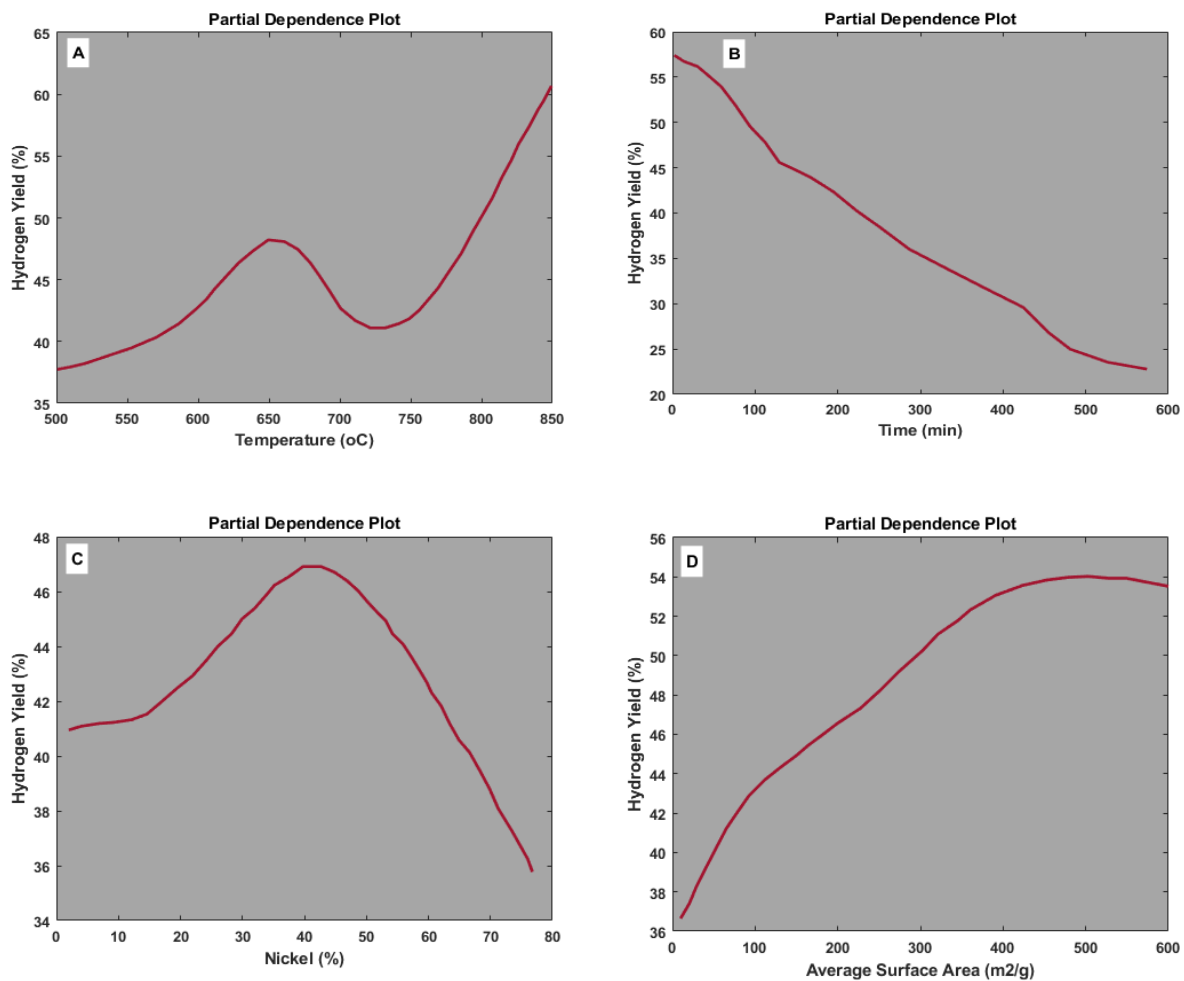
Figure 4.4B shows the GPR-based results underscore the temporal evolution of H₂ yield (%) in thermo-catalytic methane decomposition, providing a data-driven understanding of the complex interplay of factors influencing the reaction over time. It is a well-established concept in methane decomposition that during the early stages of the reaction, the catalytic sites on the surface of the catalyst material are highly active, facilitating the efficient dissociation of methane into H₂ and carbonaceous species. However, as the reaction progresses, these active sites may become gradually deactivated or fouled by the

accumulation of carbonaceous deposits, inhibiting their catalytic efficacy. The gradual decline in catalytic activity over time is often associated with the adsorption of carbon species on the catalyst surface, leading to reduced accessibility of reactive sites and hindering the methane conversion to H₂.

Figure 4.4C shows the GPR-based results depicting the effect of Nickel concentration over H₂ production. The literature shows that the concentration of nickel in catalytic systems for methane decomposition plays a pivotal role in influencing the H₂ yield (%), and its impact follows a distinctive pattern. The augmentation in H₂ yield (%), with an escalating nickel concentration can be attributed to the pivotal role that nickel plays as an active site in the methane decomposition process. As nickel concentration rises, the availability of catalytically active sites increases, promoting the efficient dissociation of methane into H₂ and carbonaceous intermediates. This heightened catalytic activity contributes to a corresponding increase in H₂ yield (%), showcasing the positive correlation between nickel concentration and the efficiency of H₂ production. However, as the nickel concentration continues to rise beyond an optimal point, a diminishing trend in H₂ yield (%) becomes apparent. This phenomenon is generally linked to the development of free Nickel inside catalyst causing catalyst deactivation by the formation of coke or carbonaceous deposits on the catalyst surface.

Figure 4.4D depicts the relationship between surface area and H₂ yield (%). The results indicate a discernible relationship between catalyst surface area and the production of H₂, revealing a notable trend. It can be postulated that initially, as the catalyst surface area is augmented, there is a corresponding increase in H₂ yield (%). The catalyst surface area serves as the interface where methane molecules interact with active sites, facilitating their dissociation into H₂ and carbonaceous intermediates. Up to a surface area of 400 m²/g, the heightened availability of active sites contributes to an enhanced catalytic activity, leading to an increased H₂ yield (%). However, beyond the critical threshold of 400 m²/g, a plateau in H₂ yield (%) is observed despite further increases in surface area. Several factors may contribute to this observed phenomenon. At higher surface areas, the catalytic sites may become saturated, limiting the additional benefit gained from increased surface exposure. Additionally, factors such as mass transport limitations or changes in the catalyst structure may become more prominent, counteracting the positive effects of increased surface area on catalytic activity. In Figure 4E, the plot illustrates the H₂ yield of various catalysts. The plot

reveals that different catalysts exhibit varying H₂ yields due to differences in catalyst material. The highest H₂ concentration was achieved by NiFe/Al₂O₃. In the plot, only the peaks and valleys of this catalyst are labeled. Figure 4.4F illustrates the impact of catalyst structure on H₂ yield, featuring eight different structures mentioned in Table 2 of the supplementary files. The plot demonstrates slight variations in H₂ yield across these structures, with the highest H₂ yield achieved by the mesoporous structure.



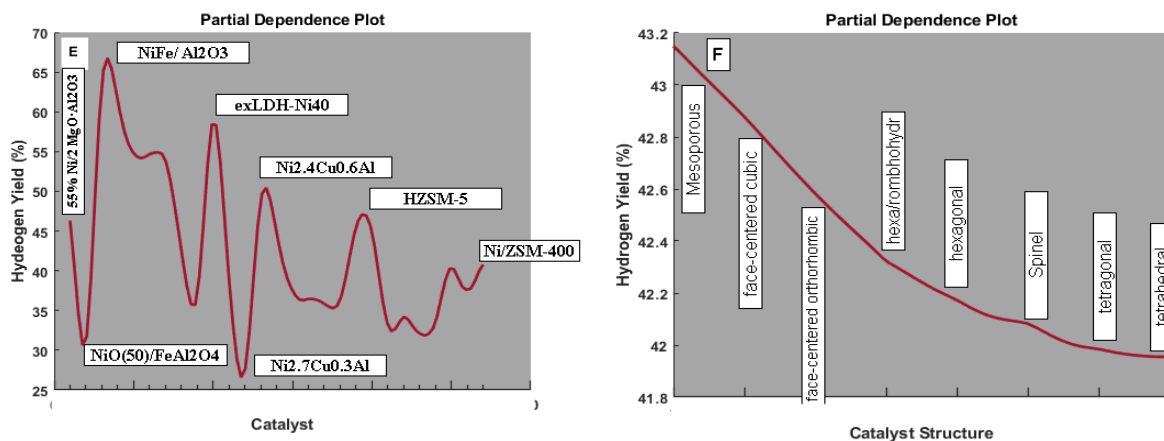


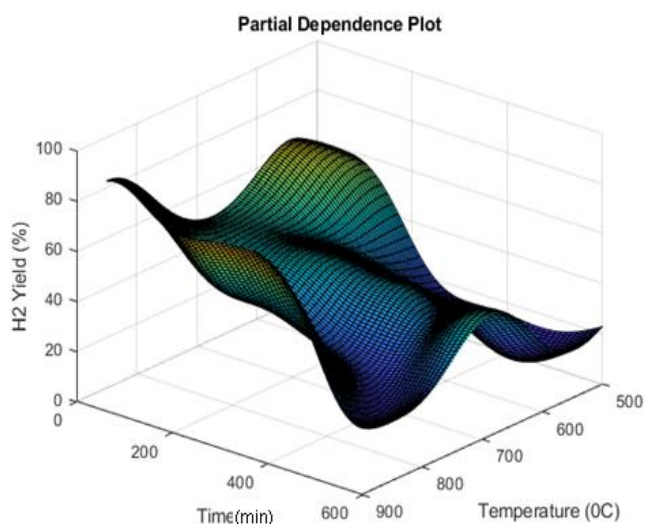
Figure 4.4: 2D Partial Dependence plots of H₂ yield (%)

Figures 4.5 (A-C) show 3D plots generated by GPR and clarify the combine effects of various parameters over H₂ production by methane decomposition. Figure 4.5A shows a trend between on stream time, temperature, and H₂ yield (%). The trend depicts overall on-stream activity of catalyst, and the results show catalytic activity of above 80% H₂ yield for 500 minutes at an operating temperature of 700°C. The observed on-stream stability under varying temperatures is crucial for the practical application of the catalytic process. It informs the design and optimization of the operating conditions to achieve the desired balance between high activity and stability. Additionally, it contributes to the overall efficiency and reliability of the catalytic system in industrial settings. The indication that on-stream stability decreases when the temperature is either increased or decreased implies a sensitivity of the catalyst to variations in thermal conditions. The reduction in stability at higher temperatures could be due to factors such as catalyst deactivation, the formation of undesirable by-products, or changes in the catalyst's selectivity under elevated thermal conditions. Similarly, if on-stream stability decreases when the temperature is decreased from 700°C, it implies that the reduction in stability could be attributed to issues such as sluggish kinetics or a shift in the predominant reaction pathways that affect the overall H₂ yield (%).

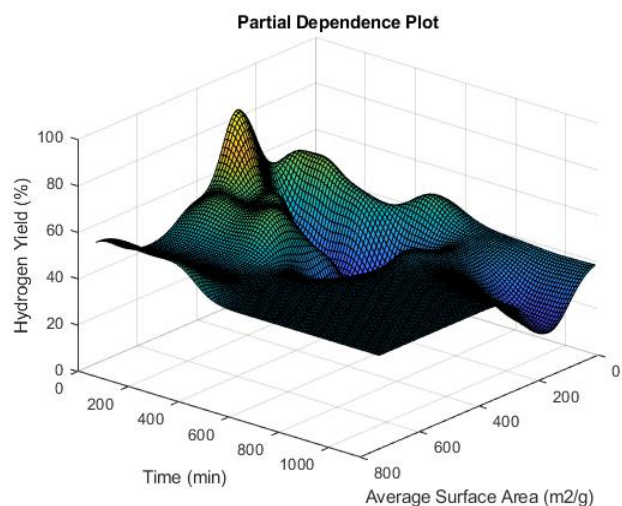
Figure 4.5 (b) depicts an interplay between time, average surface area, and H₂ yield (%). The trend reveals a compelling scenario of a catalyst exhibiting robust stability for a duration of 400 minutes, particularly at high surface areas of above 300 m²/g. This observation indicates the catalyst's resistance to deactivation mechanisms, such as carbon deposition or sintering, over an extended timeframe. The high surface area signifies a wealth of active sites available for the adsorption and decomposition of methane molecules, leading to enhanced H₂

production efficiency. The trend further elaborates the controlling factor of time in methane decomposition. At any given surface area, the H₂ yield (%) ultimately drops with the progression of time. Such insights are crucial for optimizing reaction conditions and catalyst design, paving the way for the development of more effective and durable catalysts in the realm of industrial H₂ production. Figure 4.5 (c) shows a 3D plot integrating time, H₂ yield (%), and different catalyst ranging from 55% Ni/2 MgO·Al₂O₃ to Ni/TiO₂(50%)Al₂O₃ the rest of the catalysts can be found in supplementary files. The plot reveals variations in H₂ concentration among different catalysts, yet they all follow the same trend, highlighting a notable phenomenon where a catalyst demonstrates commendable stability over longer durations. The observed stability suggests that the catalysts effectively mitigate deactivation mechanisms during the catalytic process. Scientifically, this underscores the catalyst material's role in sustaining active sites, minimizing catalyst degradation, and ensuring prolonged efficiency in converting methane to H₂. The insights gleaned from the 3D plot are crucial for catalyst design and process optimization, guiding the development of robust catalyst-support systems for enhanced and sustainable H₂ production from methane decomposition.

A



B



C

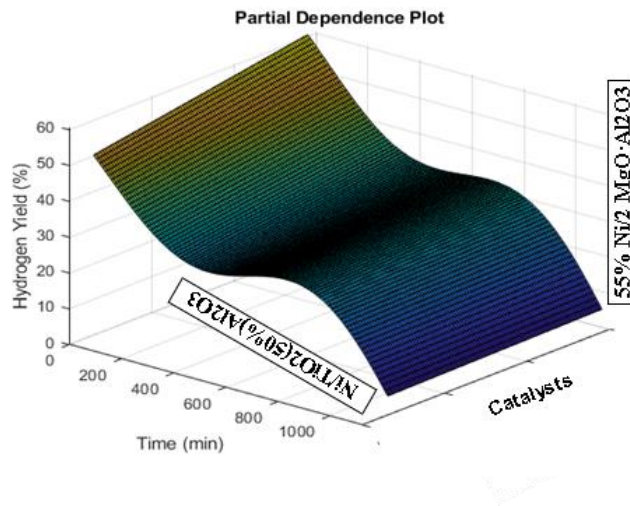


Figure 4.5: Partial dependence plots for Hydrogen Yield (%)

4.6 Graphical User Interface (GUI)

In this section, the GPR-GA methodology was employed to construct a Graphical User Interface (GUI) for forecasting H₂ yield (%) in catalytic methane decomposition. The developed GUI is visually presented in Figure 4.6. Within Figure 4.6, users have the capability to input specific parameters, including temperature (°C), time (min), and Nickel (%), and average surface area (m²/g). Additionally, users can choose catalyst type from a dropdown menu. Once all input values are provided, users can initiate the prediction process by clicking the "Predict" button. The interface enables a streamlined prediction of hydrogen yield (%), utilizing the GPR-GA model. To enhance user experience, a "Clear" button has been incorporated, allowing users to reset all input values to zero with a single click. This user-friendly design facilitates efficient and intuitive interaction with the developed GUI, enhancing the usability of the predictive model for catalytic methane decomposition.

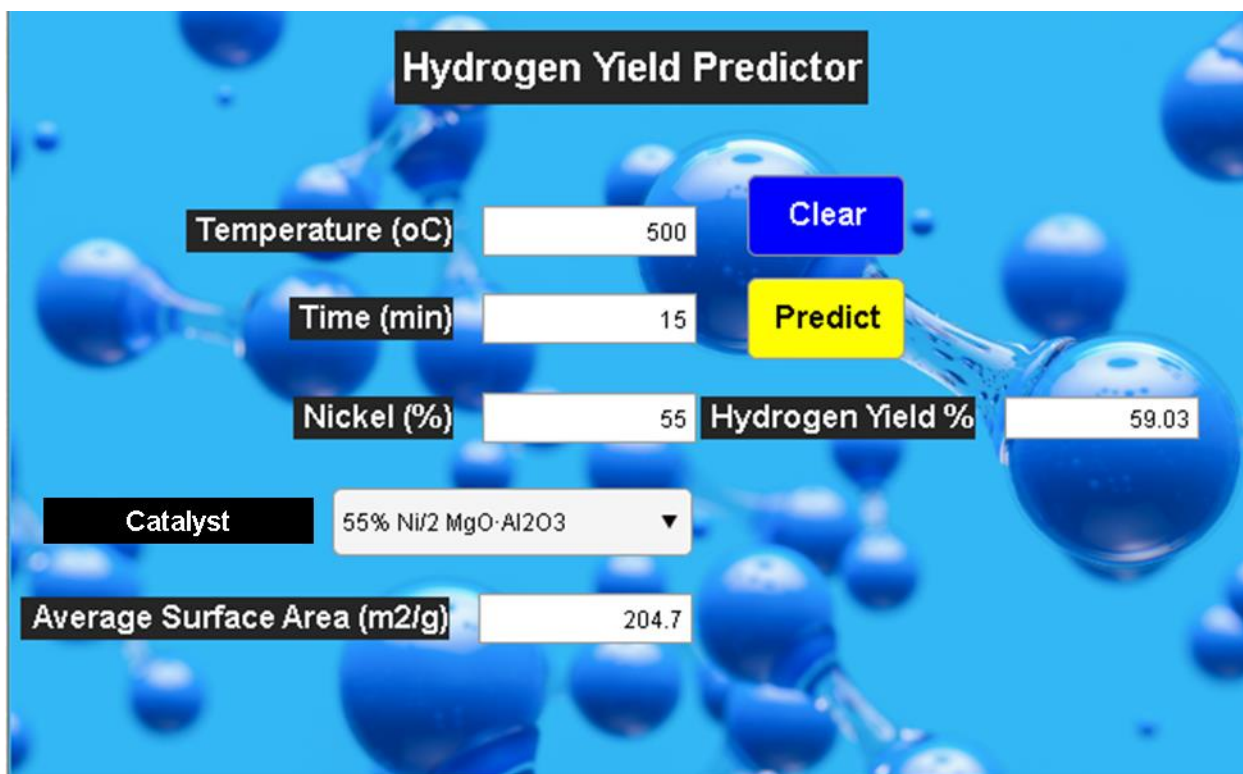


Figure 4.6: Graphical User Interface (GUI) for Hydrogen Yield (%) Prediction

4.7 Experimental Validation

Table 4.5 presents a comparative analysis of actual experimental hydrogen yields and those predicted through Gaussian Process Regression with Genetic Algorithm (GPR-GA) for three distinct scenarios. In the first instance (Sr No 1), the experimental yield was measured at 59%, closely aligned with the GPR-GA predicted yield of 59.03%. Similarly, in the second case (Sr No 2), the experimental yield reached 86%, with the GPR-GA model accurately forecasting a yield of 86.02%. The third scenario (Sr No 3) involved an experimental yield of 90%, remarkably consistent with the GPR-GA predicted yield of 90.01%. The close agreement between the experimental and predicted values across these instances demonstrates the efficacy of the GPR-GA methodology in accurately forecasting hydrogen yields in catalytic methane decomposition.

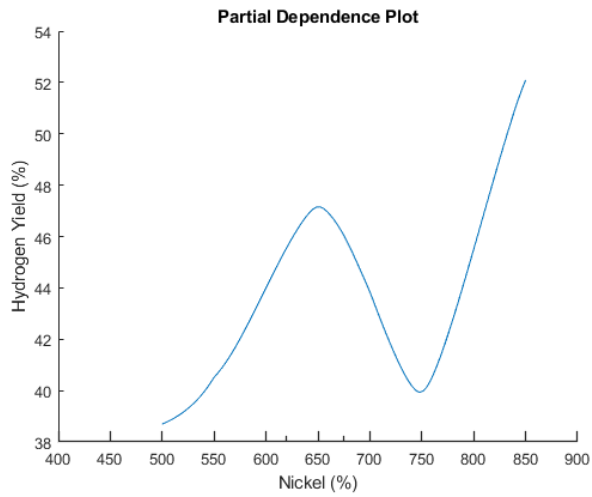
Table 4.5: Performance Table of Graphical User Interface for Hydrogen Yield Prediction

Sr No	Experimental Yield	GPR Predicted Yield	Reference
1	59	59.03	[29]
2	86	86.02	[30]
3	90	90.01	[31]

In a previous study [31], experiments were conducted using exLDH catalysts with different nickel compositions (Ni65, Ni15, Ni40) in a fixed bed reactor to assess the catalyst's performance in hydrogen production through methane decomposition. A comprehensive analysis was performed, comparing the effects of temperature and time. Here, we will compare the results of our study with the experimental study.

Regarding the impact of temperature as shown in figure 4.7, our machine learning results demonstrated that hydrogen concentration tends to increase at lower temperatures. However, as the temperature rises, there is a gradual decline in H₂ concentration, followed by a subsequent gradual increase. This trend aligns with the findings from the experimental study. Further insights into the reasons behind this observed variation in hydrogen concentration concerning temperature are provided in Section 4.5 of our study.

A



B

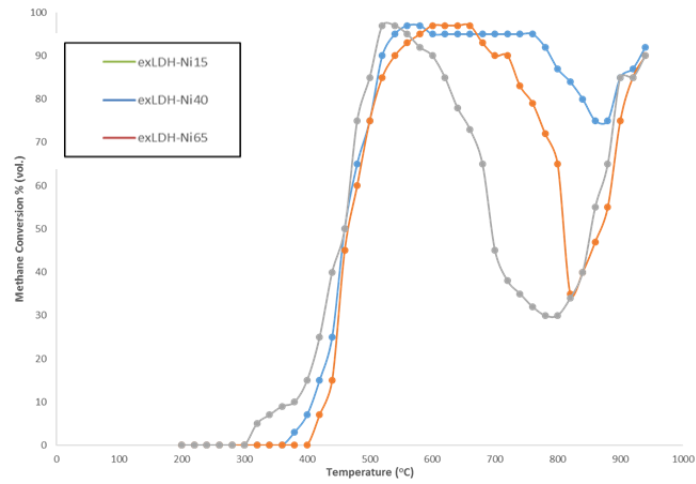


Figure 4.7: Comparison of the influence of temperature on A) ML study and B) experimental study.

Figure 4.8 illustrates a comparison of the time parameter between the M) approach employed in this study and the experimental study [31]. The observed trend in both cases indicates that with an increase in time, there is a decrease in the concentration of H₂. This phenomenon is attributed to the limited availability of active sites over time in the catalytic process.

A

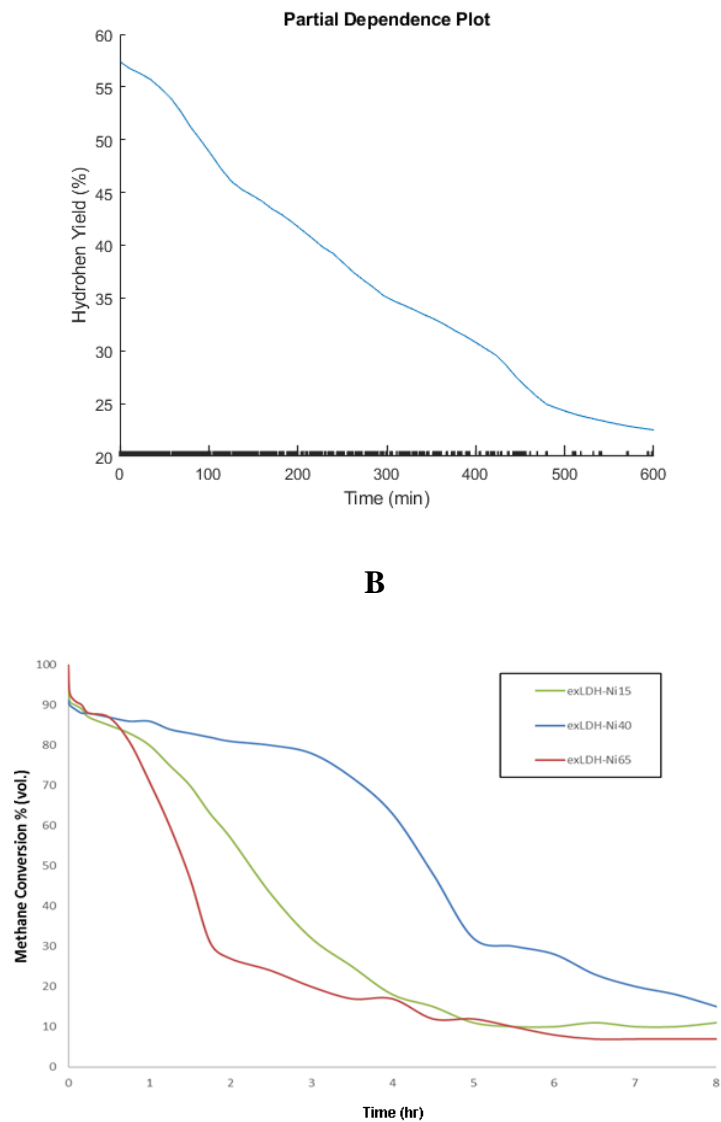


Figure 4.8: Comparison of the influence of temperature on A) ML study and B) experimental study.

The comparison of the most significant input parameters, temperature, and time, along with the use of a GUI for predicting H₂ yield, highlights the potential of machine learning in optimizing and predicting H₂ yield and associated parameters to achieve desired results.

CHAPTER 05: CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions

In conclusion, this study aimed to predict the H₂ yield (%) from catalytic methane decomposition using various machine learning (ML) models. A two-step approach was followed, incorporating feature selection and hyperparameter tuning through GA optimization. The performance of the ML models was evaluated using the R² and RMSE. The results demonstrated satisfactory performance across all models. The ET model achieved high R² values of 0.929 (training) and 0.933 (testing), with low RMSE values. GPR exhibited excellent performance, achieving a perfect training R² of 1.00 and low RMSE of 0.00026, while also performing well in testing with R² and RMSE of 0.996 and 0.063 respectively. RT showed satisfactory results, and SVM demonstrated a strong linear relationship, although with higher RMSE values. ANN achieved a high R², indicating its effectiveness in predicting H₂ yield (%). PDPs were utilized to analyze the effect of input variables on H₂ yield (%). Temperature showed a positive effect, aligning with previous literature. Time had an inverse relationship with H₂ yield (%), while Various catalysts exhibited distinct behaviors. The BET surface area demonstrated a direct linear relationship with H₂ yield (%). These visualizations provided valuable insights into the behavior and predictive capabilities of the ML models. Additionally, a user-friendly graphical user interface (GUI) was created based on the GPR-GA model, enabling users to make precise predictions regarding the H₂ yield (%). Overall, this study successfully predicted H₂ yield (%) using ML models and identified the influential factors affecting H₂ production. The findings contribute to the understanding of catalytic methane decomposition and offer insights for process optimization in H₂ production. The capability to precisely forecast H₂ yield (%) through ML models not only represents a significant leap forward in H₂ production technologies but also contributes to sustainability goals. By leveraging ML, we can enhance efficiency and optimize processes, thereby promoting cleaner production methods. This advancement holds the potential to address the escalating global energy demand while concurrently addressing environmental concerns and fostering sustainable practices in H₂ production.

5.2 Future Directions

- **Integration of Advanced Machine Learning Techniques:** Explore the potential of deep learning architectures, such as neural networks and convolutional neural networks, to improve the accuracy and robustness of hydrogen yield predictions in catalytic methane decomposition. Investigate the application of reinforcement learning algorithms for dynamic optimization of CDM processes.
- **Incorporation of Multi-scale Modeling:** Develop multi-scale modeling approaches that combine macroscopic kinetic models with atomistic-level simulations to capture the complex interactions between catalyst structures, reaction kinetics, and fluid dynamics in CDM. This integration could provide a more comprehensive understanding of CDM mechanisms and facilitate the design of highly efficient catalysts.
- **Experimental Validation and Data Acquisition:** Conduct experimental studies to validate the predictions of machine learning models under real-world operating conditions. Additionally, focus on expanding the dataset with diverse catalyst compositions, process parameters, and reactor configurations to enhance the model's predictive capabilities and generalizability.
- **Optimization of Catalyst Design:** Utilize machine learning-based optimization algorithms, such as genetic algorithms and Bayesian optimization, to systematically explore the vast design space of catalyst materials and identify novel compositions with superior performance in catalytic methane decomposition. Incorporate mechanistic insights from computational modeling to guide the search for optimal catalyst formulations.
- **Scale-Up and Techno-Economic Analysis:** Scale up the optimized CDM processes from laboratory-scale experiments to industrial production levels. Perform techno-economic analyses to assess the feasibility and economic viability of implementing CDM-based hydrogen production technologies on a commercial scale. Consider factors such as capital and operating costs, energy efficiency, and environmental impact to guide investment decisions and technology deployment strategies.
- **Exploration of Co-catalyst Systems:** Investigate the synergistic effects of incorporating multiple catalyst components or co-catalysts in CDM systems to enhance hydrogen production rates, selectivity, and catalyst stability. Explore innovative approaches, such as hybrid catalysts and composite materials, to leverage

the complementary properties of different catalytic species and maximize overall process performance.

- **Cross-disciplinary Collaboration and Knowledge Transfer:** Foster collaboration between researchers from diverse disciplines, including chemistry, chemical engineering, materials science, and data science, to leverage their expertise and accelerate progress in CDM research. Facilitate knowledge transfer and exchange of best practices through interdisciplinary workshops, conferences, and collaborative research projects to address complex challenges in sustainable hydrogen production.

REFERENCES

- [1]. Al-Mubaddel, F., et al., *H₂ production from catalytic methane decomposition using Fe/x-ZrO₂ and Fe-Ni/(x-ZrO₂)(x= 0, La₂O₃, WO₃) catalysts*. *Catalysts*, 2020. **10**(7): p. 793.
- [2]. Qian, J.X., et al., *Methane decomposition to produce CO_x-free hydrogen and nano-carbon over metal catalysts: A review*. *International Journal of Hydrogen Energy*, 2020. **45**(15): p. 7981-8001.
- [3]. Fidalgo, B. and J.Á. Menendez, *Carbon materials as catalysts for decomposition and CO₂ reforming of methane: a review*. *Chinese journal of catalysis*, 2011. **32**(1-2): p. 207-216.
- [4]. Li, Y., D. Li, and G. Wang, *Methane decomposition to CO_x-free hydrogen and nano-carbon material on group 8–10 base metal catalysts: a review*. *Catalysis today*, 2011. **162**(1): p. 1-48.
- [5]. Ayodele, B.V., et al., *Artificial intelligence modelling approach for the prediction of CO-rich hydrogen production rate from methane dry reforming*. *Catalysts*, 2019. **9**(9): p. 738.
- [6]. Kathiraser, Y., et al., *Kinetic and mechanistic aspects for CO₂ reforming of methane over Ni based catalysts*. *Chemical Engineering Journal*, 2015. **278**: p. 62-78.
- [7]. Arce-Medina, E. and J.I. Paz-Paredes, *Artificial neural network modeling techniques applied to the hydrodesulfurization process*. *Mathematical and Computer Modelling*, 2009. **49**(1-2): p. 207-214.
- [8]. Puig-Arnavat, M. and J. Bruno, *Artificial Neural Networks for Thermochemical Conversion of Biomass*. *Recent Adv. Thermo-Chem. Convers.* 2015, Biomass.
- [9]. Ghasemzadeh, K., et al., *Hydrogen production by a PdAg membrane reactor during glycerol steam reforming: ANN modeling study*. *International Journal of Hydrogen Energy*, 2018. **43**(15): p. 7722-7730.

- [10]. George, J., P. Arun, and C. Muraleedharan, *Assessment of producer gas composition in air gasification of biomass using artificial neural network model*. International Journal of Hydrogen Energy, 2018. **43**(20): p. 9558-9568.
- [11]. Basile, A., et al., *Water gas shift reaction in membrane reactors: Theoretical investigation by artificial neural networks model and experimental validation*. international journal of hydrogen energy, 2015. **40**(17): p. 5897-5906.
- [12]. Shahbaz, M., et al., *Artificial neural network approach for the steam gasification of palm oil waste using bottom ash and CaO*. Renewable Energy, 2019. **132**: p. 243-254.
- [13]. Nasr, N., et al., *Application of artificial neural networks for modeling of biohydrogen production*. International journal of hydrogen energy, 2013. **38**(8): p. 3189-3195.
- [14]. Ayodele, B.V. and C.K. Cheng, *Modelling and optimization of syngas production from methane dry reforming over ceria-supported cobalt catalyst using artificial neural networks and Box–Behnken design*. Journal of Industrial and Engineering Chemistry, 2015. **32**: p. 246-258.
- [15]. Haq, Z.U., et al., *Hydrogen production optimization from sewage sludge supercritical gasification process using machine learning methods integrated with genetic algorithm*. Chemical Engineering Research and Design, 2022. **184**: p. 614-626.
- [16]. Kotsiantis, S.B., I. Zaharakis, and P. Pintelas, *Supervised machine learning: A review of classification techniques*. Emerging artificial intelligence applications in computer engineering, 2007. **160**(1): p. 3-24.
- [17]. Pino-Mejías, R., et al., *Comparison of linear regression and artificial neural networks models to predict heating and cooling energy demand, energy consumption and CO2 emissions*. Energy, 2017. **118**: p. 24-36.
- [18]. Ullah, H., et al., *Optimization based comparative study of machine learning methods for the prediction of bio-oil produced from microalgae via pyrolysis*. Journal of Analytical and Applied Pyrolysis, 2023: p. 105879.

- [19]. Roberge, V., M. Tarbouchi, and G. Labonté, *Comparison of parallel genetic algorithm and particle swarm optimization for real-time UAV path planning*. IEEE Transactions on industrial informatics, 2012. **9**(1): p. 132-141.
- [20]. Oliveira, A.L., et al., *GA-based method for feature selection and parameters optimization for machine learning regression applied to software effort estimation*. information and Software Technology, 2010. **52**(11): p. 1155-1166.
- [21]. Tabassum, M. and K. Mathew, *A genetic algorithm analysis towards optimization solutions*. International Journal of Digital Information and Wireless Communications (IJDIWC), 2014. **4**(1): p. 124-142.
- [22]. Kumar, S., S. Jain, and H. Sharma, *Genetic algorithms*, in *Advances in swarm intelligence for optimizing problems in computer science*. 2018, Chapman and Hall/CRC. p. 27-52.
- [23]. Albadr, M.A., et al., *Genetic algorithm based on natural selection theory for optimization problems*. Symmetry, 2020. **12**(11): p. 1758.
- [24]. García-Martínez, C., et al., *Global and local real-coded genetic algorithms based on parent-centric crossover operators*. European journal of operational research, 2008. **185**(3): p. 1088-1113.
- [25]. Li, Z., et al., *Influence of AlCl₃ and oxidant catalysts on hydrogen production from the supercritical water gasification of dewatered sewage sludge and model compounds*. International Journal of Hydrogen Energy, 2021. **46**(61): p. 31262-31274.
- [26]. Ladha, L. and T. Deepa, *Feature selection methods and algorithms*. International journal on computer science and engineering, 2011. **3**(5): p. 1787-1797.
- [27]. Vafaie, H. and K.A. De Jong. *Genetic Algorithms as a Tool for Feature Selection in Machine Learning*. in *ICTAI*. 1992. Citeseer.
- [28]. Wu, J., et al., *Hyperparameter optimization for machine learning models based on Bayesian optimization*. Journal of Electronic Science and Technology, 2019. **17**(1): p. 26-40.

- [29]. Rastegarpanah, A., et al., *Thermocatalytic decomposition of methane over mesoporous Ni/xMgO·Al₂O₃ nanocatalysts*. International Journal of Hydrogen Energy, 2018. **43**(32): p. 15112-15123.
- [30]. Ergazieva, G.E., et al., *Catalytic Decomposition of Methane to Hydrogen over Al₂O₃ Supported Mono- and Bimetallic Catalysts*. Bulletin of Chemical Reaction Engineering & Catalysis, 2022. **17**(1): p. 1-12.
- [31]. Sikander, U., et al., *Tailored hydrotalcite-based Mg-Ni-Al catalyst for hydrogen production via methane decomposition: Effect of nickel concentration and spinel-like structures*. International Journal of Hydrogen Energy, 2019. **44**(28): p. 14424-14433.