Carbon Capture Prediction by Metallic Organic Frameworks using Machine Learning

Project-III (CH57003) report submitted to

Indian Institute of Technology Kharagpur

in partial fulfilment for the award of the degree of

Master of Technology

in

Chemical Engineering

by

Luqmaan Ahmad

(20CH30015)

Under the supervision of

Professor Dr. Sudipto Chakraborty

Department of Chemical Engineering

Indian Institute of Technology Kharagpur

Autumn Semester, 2024-25

December 28, 2024

DECLARATION

I certify that

(a) The work contained in this report has been done by me under the guidance of my supervisor.

(b) The work has not been submitted to any other Institute for any degree or diploma.

(c) I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

(d) Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Date: December 28, 2024 (Luqmaan Ahmad)

Place: Kharagpur (20CH30015)

DEPARTMENT OF CHEMICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

KHARAGPUR - 721302, INDIA

CERTIFICATE

This is to certify that the project report entitled "**Carbon Capture Prediction by Metallic Organic Frameworks using Machine Learning**" submitted by **Luqmaan Ahmad** (Roll No. 20CH30015) to Indian Institute of Technology Kharagpur towards partial fulfilment of requirements for the award of degree of Master of Technology in Chemical Engineering is a record of bona fide work carried out by him under my supervision and guidance during Autumn Semester, 2024-25.

Date: December 28, 2024
Place: Kharagpur

Professor Dr. Sudipto Chakraborty
Department of Chemical Engineering
Indian Institute of Technology Kharagpur
Kharagpur - 721302, India

Abstract

Name of the student: **Luqmaan Ahmad** Roll No: **20CH30015**

Degree for which submitted: **Master of Technology**

Department: **Department of Chemical Engineering**

Thesis title: **Carbon Capture Prediction by Metallic Organic Frameworks using Machine Learning**

Thesis supervisor: **Professor Dr. Sudipto Chakraborty**

Month and year of thesis submission: **December 28, 2024**

This thesis aims to explore the application of machine learning in predicting the CO adsorption capacity of metal-organic frameworks (MOFs) to advance carbon capture technologies. A dataset of 3,800 entries, comprising textural and process properties, was compiled from experimental studies and literature. Predictive models, including Random Forest, Gradient Boosting, and Artificial Neural Networks, were evaluated, with Random Forest emerging as the best-performing model.

Key findings highlight the importance of features like BET surface area and pore volume in determining adsorption efficiency. However, the absence of chemical properties such as acidic/basic sites underscores the need for more comprehensive datasets. Future work will focus on incorporating additional features and employing explainable AI to further enhance model interpretability and guide the design of high-performance MOFs for sustainable carbon capture.

Acknowledgements

I express my deepest gratitude to my supervisor, Prof. Sudipto Chakraborty, for his invaluable guidance, insightful advice, and unwavering support throughout the course of this thesis. His expertise and encouragement have been instrumental in shaping the direction of this project and ensuring its successful completion. I am also immensely thankful to my PhD advisor, Mr. Deepu, whose mentorship and constructive feedback have significantly contributed to refining my work. Their patience and belief in my abilities have been a constant source of motivation.

My deepest thanks go to my parents for their unwavering love, support, and encouragement throughout this academic journey. Their belief in my potential and sacrifices have been a constant source of strength, allowing me to pursue my goals with determination and confidence. Their endless patience and motivation have been the cornerstone of my success.

I am also profoundly grateful to my fellow batchmate, whose collaborative efforts and insightful contributions have greatly enriched this project. Their willingness to discuss ideas, troubleshoot challenges, and share perspectives has made this experience both rewarding and enjoyable. The camaraderie and shared commitment to excellence have been invaluable in completing this work.

This project is the culmination of collective efforts, and I am sincerely thankful to everyone who has supported me along the way. From peers who offered advice to faculty members who provided valuable input, each contribution has played a vital role in this academic endeavor. To all those who have inspired, encouraged, and assisted me, I offer my heartfelt gratitude.

This journey has been an incredible learning experience, and I am truly blessed to have had the support and guidance of such wonderful individuals. Thank you for being an integral part of this achievement.

Contents

Chapter 1

Introduction

The increasing concentration of carbon dioxide (CO) in the atmosphere has become one of the leading causes of global climate change. This rise is linked to industrialization, energy production, and other anthropogenic activities that release vast amounts of CO into the environment. The consequences of unchecked CO emissions include global warming, extreme weather events, rising sea levels, and significant ecological imbalances. To mitigate these impacts, carbon capture technologies have emerged as critical tools for reducing greenhouse gas emissions and slowing the progression of climate change.

Carbon capture is a process of trapping CO at its source, such as power plants and heavy industrial facilities, before it is released into the atmosphere. This approach targets major emitters, contributing significantly to reducing global carbon emissions. The development and deployment of efficient carbon capture solutions are pivotal for achieving climate goals and transitioning toward a sustainable future.

Why Use Metal-Organic Frameworks (MOFs) for Carbon Capture?

Metal-organic frameworks (MOFs) have gained considerable attention as one of the most effective materials for carbon capture. MOFs are highly porous, crystalline

materials formed by linking metal ions with organic ligands. Their unique properties, such as customizable pore sizes, high surface areas, and exceptional chemical stability, make them ideal for adsorbing CO. MOFs outperform many traditional materials due to their ability to be tailored at the molecular level to optimize performance.

The structural versatility of MOFs allows for the incorporation of functional sites that enhance their selectivity and capacity for CO adsorption. These materials can be engineered to preferentially adsorb CO over other gases, making them highly efficient in separating CO from industrial emissions. Additionally, MOFs exhibit strong potential for reusability, reducing operational costs and environmental impact over time.

The Role of MOFs in Mitigating CO Emissions

MOFs play a crucial role in mitigating CO emissions through both adsorption and conversion processes. In adsorption, MOFs capture CO molecules within their porous structures, effectively reducing emissions at the source. Some MOFs, such as those with open metal sites, demonstrate exceptional CO uptake even under low-pressure conditions.

Beyond capture, MOFs also contribute to the conversion of CO into valuable chemicals such as methanol, formic acid, and urea. This dual capability supports a circular carbon economy, turning a harmful greenhouse gas into useful products. By enabling both capture and utilization, MOFs offer a comprehensive approach to addressing carbon emissions.

The Need for Sustainable Carbon Capture Solutions

As global challenges related to climate change intensify, the demand for innovative and efficient carbon capture solutions becomes increasingly urgent. MOFs represent a transformative advancement in this domain, providing highly effective and adaptable materials for capturing and utilizing CO. Their unique characteristics,

combined with ongoing research and development, position MOFs as a cornerstone technology in the fight against climate change. Investing in MOF-based carbon capture systems can drive significant progress toward reducing emissions and achieving long-term sustainability goals

Chapter 2

Literature Review

2.1 Introduction

In this chapter, we explore previous works related to carbon capture technologies, focusing on the utilization of metal-organic frameworks (MOFs) and the integration of machine learning for optimizing adsorption-based carbon capture.

2.2 Carbon Capture and Conversion Using Metal-Organic Frameworks

This paper provides a comprehensive overview of the role of metal-organic frameworks (MOFs) in carbon capture and conversion technologies. MOFs are presented as advanced materials with unparalleled potential due to their high surface area, tunable porosity, and chemical stability. The authors detail how MOFs can selectively capture CO molecules even under low partial pressures, making them highly effective for industrial applications.

The study delves into the structural features of MOFs that contribute to their efficiency, such as the incorporation of functional groups and the presence of open metal sites. Specific examples, like Mg-MOF-74, are highlighted for their exceptional CO adsorption capacities and strong binding interactions with CO molecules. Furthermore, the paper explores the dual functionality of MOFs in capturing CO and converting it into valuable products like methanol and urea, thus integrating capture and utilization into a circular carbon economy.

Additionally, the authors examine the challenges in scaling up MOFs for practical applications, including their regeneration and reusability. Various regeneration techniques, such as pressure swing and thermal swing processes, are discussed as methods to reduce operational costs. The paper concludes with insights into the future directions for enhancing MOF performance through computational design and experimental innovations.

2.3 Explainable Machine Learning for Carbon Capture Optimization

This paper emphasizes the growing role of explainable machine learning (ML) models in optimizing carbon capture technologies. Unlike traditional "black-box" ML approaches, explainable ML provides interpretability and insights into the factors influencing adsorption processes. The study focuses on the use of techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) to identify critical variables affecting CO adsorption performance.

The authors present a framework that combines experimental data with ML algorithms to predict adsorption capacities and selectivity of materials such as MOFs. By incorporating interpretability, the study bridges the gap between predictive modeling

and actionable insights, enabling researchers to understand the impact of features like surface area, pore volume, and adsorption pressure on performance.

A key contribution of the paper is its use of tree-based models, including Random Forest and Gradient Boosting, for predicting adsorption metrics. These models were found to achieve high accuracy while maintaining transparency in their decision-making processes. The study concludes that explainable ML can accelerate the identification of optimal materials for carbon capture while providing a deeper understanding of the underlying adsorption mechanisms.

2.4 Machine Learning-Assisted Carbon Capture Prediction

This paper focuses on leveraging machine learning (ML) to streamline the design and selection of materials for adsorption-based carbon capture. The authors address the challenge of scattered data on adsorbent properties by consolidating and preprocessing a comprehensive database. This database includes information on textural properties (e.g., surface area, pore volume, pore diameter) and operating conditions (e.g., temperature, pressure).

The study employs diversified ML techniques, such as Random Forest, Deep Neural Networks (DNNs), and Gradient Boosting, to classify adsorbents based on their CO adsorption capacity and selectivity. These models were used to extract critical design rules, such as the importance of large surface areas and small pore diameters for achieving high selectivity. One of the key findings was the identification of adsorption pressure as a major determinant of capacity, followed by temperature and pore volume.

Explainable AI techniques, such as SHAP, were also utilized to interpret the ML models. This helped in uncovering the relative importance of various features on

adsorption performance, providing actionable insights for material design. The paper concludes by emphasizing the need for data-driven methodologies to accelerate the discovery of efficient adsorbents while reducing the reliance on costly and time-consuming experimental methods.

Summary

These papers collectively highlight the intersection of advanced materials and machine learning in revolutionizing carbon capture technologies. The first paper focuses on the fundamental capabilities of MOFs as high-performance adsorbents, while the second and third papers emphasize the role of explainable machine learning in optimizing and accelerating material selection. Together, they provide a comprehensive understanding of the current advancements and future potential in adsorption-based carbon capture solutions.

Chapter 3

Critical Factors for MOF-Based Carbon Capture

In this chapter, we discuss the various factors that influence the carbon capture efficiency of metal-organic frameworks (MOFs). These factors can be broadly categorized into textural properties, operating conditions, and chemical properties, all of which play crucial roles in determining the adsorption capacity and selectivity of MOFs.

3.1 Textural Properties

Textural properties are physical characteristics of MOFs that directly affect their ability to adsorb CO_2. These include:

3.1.1 BET Specific Surface Area (BET)

The Brunauer-Emmett-Teller (BET) surface area is a measure of the total surface area of a material. A higher BET surface area provides more sites for CO_2 molecules

to adsorb, significantly enhancing the adsorption capacity. MOFs with large surface areas, such as MIL-101 and UiO-66, have demonstrated superior CO_2 uptake due to the increased availability of adsorption sites.

3.1.2 Pore Volume (VP)

The pore volume of a MOF refers to the total void space within its structure. A larger pore volume allows for greater gas storage capacity, enabling the MOF to accommodate more CO_2 molecules. This property is particularly important for MOFs designed for high-pressure adsorption applications.

3.1.3 Micropore Volume (VMP)

Micropores, with diameters less than 2 nm, play a critical role in selective CO_2 adsorption. The high energy density within micropores enhances CO_2 binding, especially at low pressures, making micropore volume a key factor in adsorption efficiency.

3.1.4 Pore Diameter (DP)

The pore diameter affects the diffusion of CO_2 molecules into the MOF structure. MOFs with optimal pore sizes that match the kinetic diameter of CO_2 (approximately 3.3 Å) exhibit higher adsorption efficiencies. Narrow pores improve selectivity by excluding larger gas molecules, such as N and CH.

3.2 Operating Conditions

Operating conditions, such as temperature, pressure, and CO_2 partial pressure, have a significant impact on MOF performance:

3.2.1 Temperature (T)

The adsorption of CO_2 is typically an exothermic process, meaning that lower temperatures favor adsorption. At elevated temperatures, the adsorption capacity decreases due to reduced binding strength between CO_2 molecules and the adsorption sites. Thus, MOFs designed for carbon capture are often tested and optimized for low-temperature conditions.

3.2.2 Pressure (P)

Higher pressures generally enhance CO_2 adsorption capacity as the increased gas density leads to more CO_2 molecules being captured. However, the adsorption capacity often plateaus beyond a certain pressure due to the saturation of adsorption sites.

3.2.3 CO_2 Partial Pressure (CPP)

The partial pressure of CO_2 in a gas mixture affects the selectivity and efficiency of MOFs. Higher CO_2 partial pressures increase adsorption, but for practical applications, MOFs need to perform well at low partial pressures, such as those encountered in flue gases.

3.3 Chemical Properties

The chemical composition and functionality of MOFs significantly influence their adsorption behavior:

3.3.1 Acidic and Basic Sites

Basic Sites: MOFs with basic functional groups or open metal sites exhibit enhanced CO_2 adsorption due to the interaction between the electron-deficient CO_2 molecules and the basic sites. For example, Mg-MOF-74 is highly effective in CO_2 capture because of its open Mg^2 sites.

Acidic Sites: Acidic functional groups, such as carboxylates, improve selectivity by interacting with CO_2 through hydrogen bonding or dipole-quadrupole interactions. These sites are particularly useful in environments with competing gases like N or CH.

3.3.2 Functionalization

Functionalizing MOFs with amines or other polar groups can significantly improve their affinity for CO_2. Amines, for instance, form carbamate species upon reacting with CO_2, increasing the adsorption capacity even at low pressures.

3.3.3 Hydrophobicity

Hydrophobic MOFs are more stable and effective in humid environments. They prevent the adsorption of water, which can compete with CO_2 for adsorption sites and degrade the MOF structure over time.

Summary The efficiency of MOFs in carbon capture is determined by a complex interplay of textural, operational, and chemical factors. High BET surface areas, optimized pore structures, and functionalized chemical sites enhance their adsorption capacity and selectivity. Moreover, the performance of MOFs is significantly influenced by operating conditions, underscoring the importance of tailoring MOF properties to specific industrial requirements. A thorough understanding of these factors is essential for the design and development of next-generation MOFs for efficient and sustainable carbon capture.

Chapter 4

Data Collection and Interpretation

The scattered nature of data on carbon capture materials presents a significant challenge in designing effective adsorption-based solutions. Data related to adsorption properties, material characteristics, and operating conditions is often dispersed across numerous research papers, literature reviews, and official databases. To address this, an extensive data collection process was undertaken to compile a comprehensive dataset focusing on the factors influencing carbon capture in metal-organic frameworks (MOFs) and related materials.

This chapter details the systematic approach followed to gather, standardize, and preprocess the data, the structure of the resulting dataset, and the visualizations developed to explore its characteristics. Additionally, it highlights the challenges encountered, including gaps in the availability of critical features like acidic/basic sites and functionalization.

4.1 Data Collection Methodology

Source Identification

The data was collected from multiple sources, including:

Original research papers that reported experimental results. Literature reviews summarizing existing studies. Official open-access databases dedicated to MOF data. Simulated data repositories and computational studies. A specific dataset derived from an open-source MOF repository focusing on textural and adsorption properties.

Variables of Interest The dataset was designed to capture variables critical to understanding carbon capture performance. These include:

Adsorption Properties:

CO adsorption capacity (mmol/g) and IAST CO/N selectivity. Textural Properties:

BET specific surface area (BET): Measures the available surface area for adsorption.

Pore volume (VP): Reflects the material's capacity to store gases.

Micropore volume (VMP): Highlights the contribution of small pores (¡2 nm) in adsorption.

Pore diameter (DP): Dictates the diffusion and selectivity of gases. Process Conditions:

Capture temperature (T) and pressure (P): Critical for adsorption thermodynamics.

CO partial pressure (CPP): A key parameter in selective adsorption studies.

4.2 Preprocessing and Data Standardization

Data Compilation

The extracted data underwent extensive preprocessing to ensure consistency and usability. Features were aligned across sources by:

Converting units to a common standard (e.g., pressure in bar, temperature in °C). Filtering out incomplete or inconsistent entries. Defining thresholds for extreme values to remove outliers.

Missing Features

While the dataset captured a wide range of variables, certain chemical characteristics, such as acidic/basic sites and functionalization, were not consistently reported. These properties, though critical for understanding adsorption mechanisms, were either absent or scattered across papers, limiting their inclusion in the analysis.

4.3 Dataset Overview

The final dataset consists of approximately 3,800 rows, representing individual materials or experimental conditions. Each row includes information on textural, chemical, and process properties, along with adsorption performance metrics. Below is a sample of the dataset:

Adsorbent_Name	Adsorbent ID	Chemical Formula	MOF Code	Symmetry	Cell Density [g/cm^3]	Cell Volume []	Pore Diameter [Å]	Porosity [-]	Pressure [bar]	Temperature [°C]	CO2 Partial Pressure [bar]	CO2 Adsorption Capacity [mmol/g]
[tris(Dimethylammonium) octakis(μ6-benzene-1,3...	262	(C2 H8 N1 1+)3n,n(C72 H24 Cl3 Cu12 O48 3-),9n(...	ABEMIF	Pm-3m	1.85	3640.89	6.80	0.64	1	25	0.15	0.727365
(bis(Tetrapropylammonium) hexakis(μ5-isomaleon...	263	(C12 H28 N1 1+)2n,n(C28 H6 Cu10 N14 S12 2-),0....	ABINEF	C2/c	1.27	10127.90	4.62	0.55	1	25	0.15	0.672152
(bis((μ3-Imidazole-5,6-dicarboxylato)-(4,4'-bi...	264	(C40 H28 Co3 N10 O8)n,7n(H2 O1)	ABIYIV	Pccn	1.44	4703.28	3.88	0.56	1	25	0.15	1.041453

FIGURE 4.1: Dataset Sample

The dataset captures diverse MOF properties but remains incomplete for certain chemical descriptors, necessitating further data collection in future studies.

4.4 Visualization and Data Exploration

To better understand the dataset, exploratory data analysis (EDA) techniques were employed, including:

4.4.1 Textural Properties

Histograms and box plots were used to examine the distribution of features like BET surface area, pore volume, and pore diameter. These visualizations highlighted key trends, such as:

MOFs with larger BET surface areas generally exhibit higher adsorption capacities. Micropores contribute significantly to CO selectivity due to their size compatibility with CO molecules. 4.5.2 Process Conditions Scatter plots and line graphs revealed the effects of temperature, pressure, and CO partial pressure on adsorption performance. For instance:

CO adsorption decreases at higher temperatures due to the exothermic nature of the adsorption process. Higher pressures result in saturation of adsorption sites, leading to a plateau in capacity.

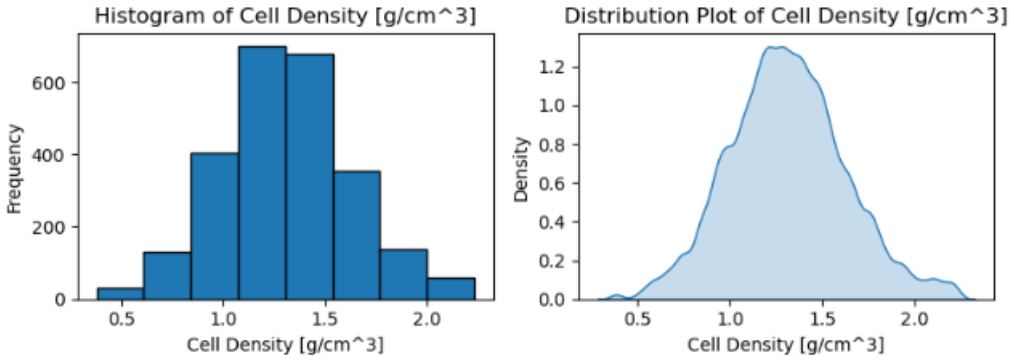

FIGURE 4.2: cell density

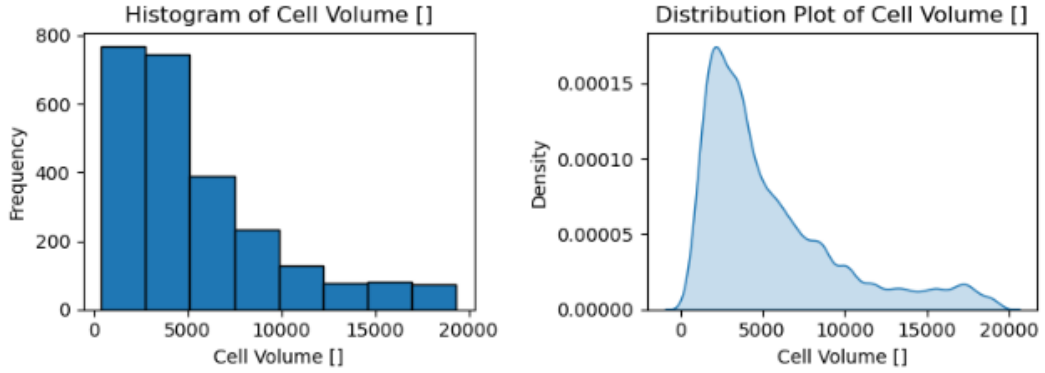

FIGURE 4.3: Cell volume

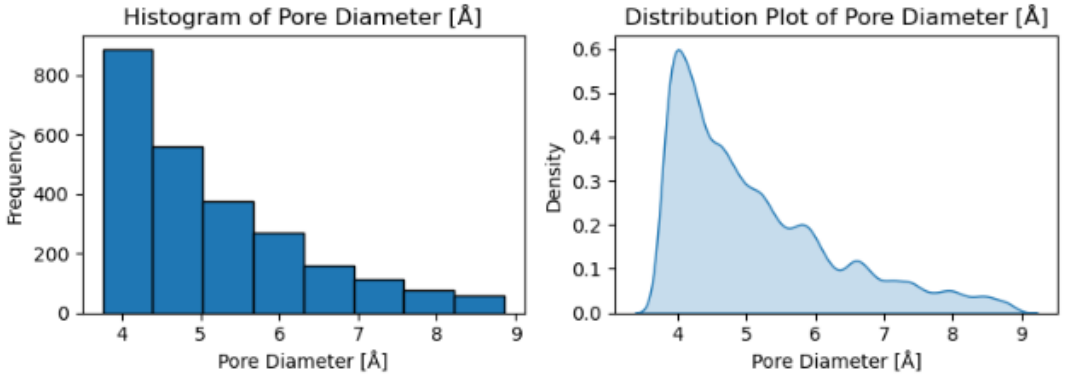

FIGURE 4.4: Pore Diameter

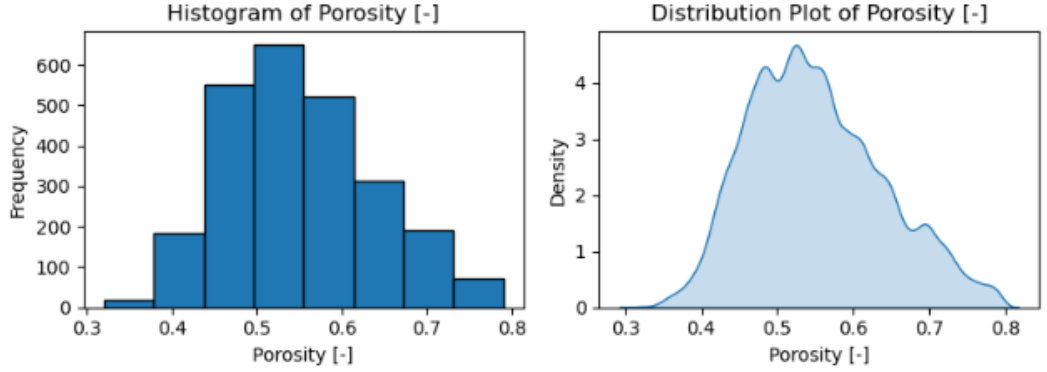

FIGURE 4.5: Porosity

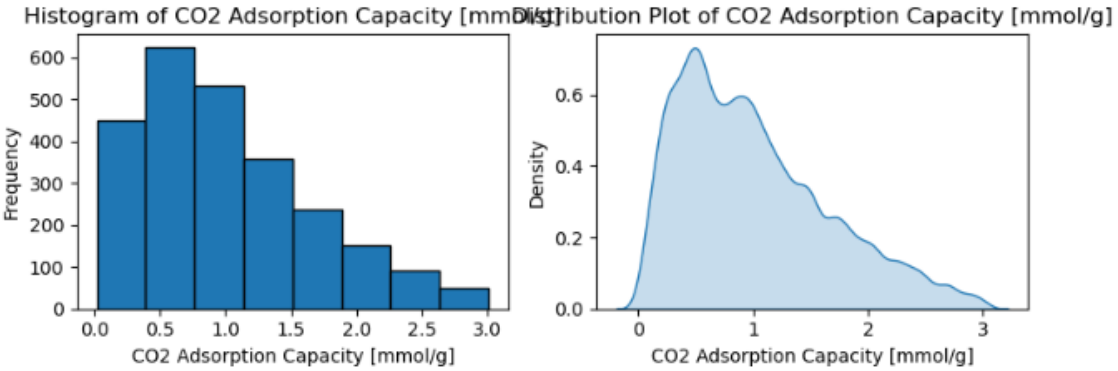

FIGURE 4.6: Adsorption capacity

4.4.2 Correlation Analysis

Correlation matrices were generated to identify relationships between features. Strong correlations were observed between:

BET surface area and CO adsorption capacity. Pore diameter and selectivity.

Two different kind of the correlation matrix are formed which are pearson and spearman matrix.

4.5 Challenges and Limitations

4.5.1 Incomplete Data

The absence of detailed information on chemical properties like acidic/basic sites and functionalization limited the scope of certain analyses. These properties are critical for designing MOFs with enhanced CO affinity and selectivity.

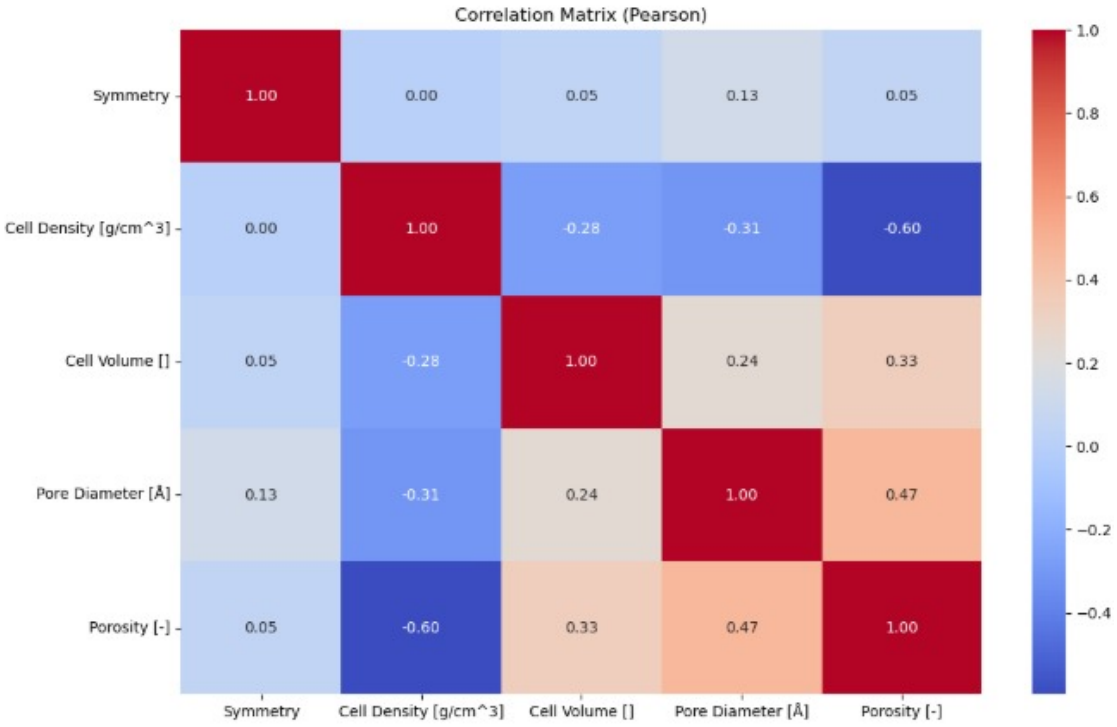

FIGURE 4.7: Pearson Correlation Matrix

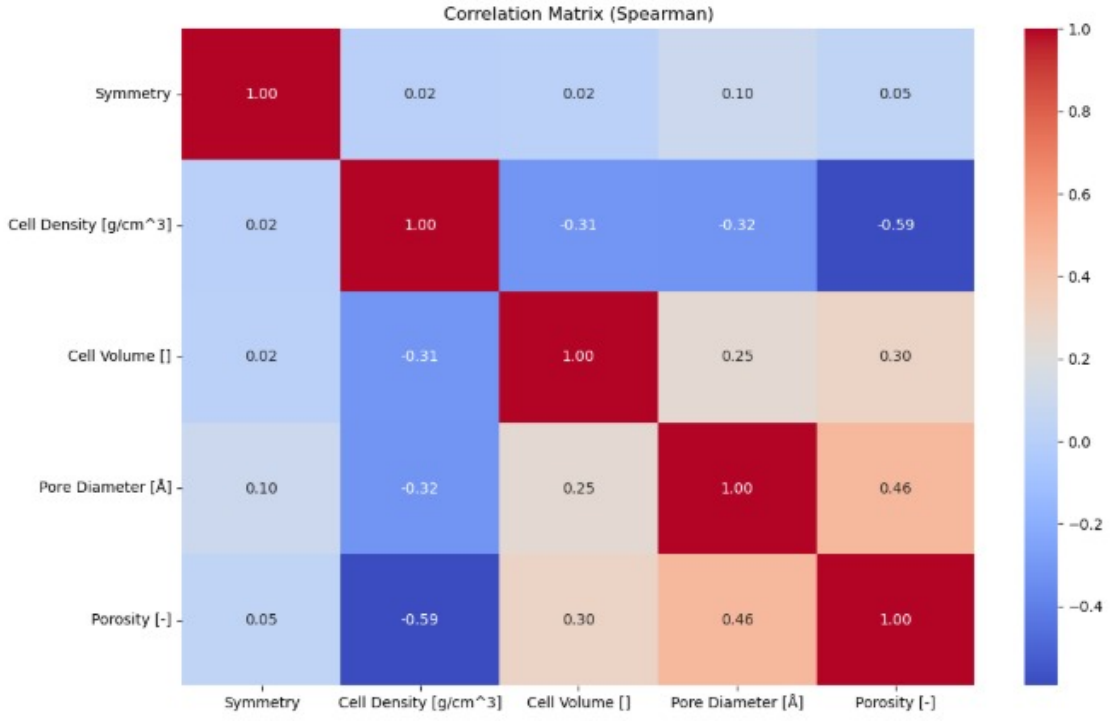

FIGURE 4.8: Spearman Correlation Matrix

4.5.2 Variability in Reporting

Differences in reporting standards across sources required significant preprocessing and standardization efforts. Discrepancies in units, formats, and experimental conditions posed challenges in aligning the data.

4.5.3 Complex Interdependencies

The interdependence of features, such as temperature and pore structure, introduces non-linear relationships that are challenging to model with conventional statistical methods.

Summary

This chapter outlines the systematic approach to collecting and preprocessing data for carbon capture analysis. The resulting dataset provides a robust foundation for modeling and prediction, capturing key textural, chemical, and process variables. However, gaps in the data, particularly related to chemical properties, highlight the need for continued data collection and integration. The next chapter delves into model building and the predictive insights derived from this dataset.

Chapter 5

Model Performance Analysis

The performance of machine learning models was evaluated to predict the CO adsorption capacity of materials based on their textural, process, and chemical features. Multiple regression models were implemented, and their performance was assessed using evaluation metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R^2 score. A comparative analysis was conducted to identify the best-performing models and their suitability for predicting adsorption capacities.

5.1 Models Implemented

Several machine learning models were implemented, ranging from simple linear models to advanced ensemble techniques. The models used include:

5.1.1 Linear Regression

: A baseline model used to establish a benchmark for performance.

5.1.2 Random Forest Regressor

: An ensemble learning technique that uses multiple decision trees to improve prediction accuracy.

5.1.3 Decision Tree Regressor

: A single-tree approach, useful for understanding how individual features contribute to predictions.

5.1.4 Support Vector Regressor (SVR)

: A kernel-based regression method capable of handling non-linear relationships.

5.1.5 K-Nearest Neighbors Regressor (KNN)

: A simple distance-based method for predicting values based on similar instances.

5.1.6 Gradient Boosting Regressor

: A sequential ensemble model that reduces prediction errors over iterations.

5.1.7 AdaBoost Regressor

: An adaptive boosting technique that emphasizes learning from previous errors.

5.2 Evaluation Metrics

The models were evaluated using the following metrics:

5.2.1 Mean Squared Error (MSE)

: Measures the average squared error between predicted and actual values. Lower values indicate better performance.

5.2.2 R^2 Score

: Indicates the proportion of variance explained by the model. Higher values represent better predictive power.

5.2.3 Train/Test Loss

: Used to analyze overfitting or underfitting behavior in the models.

5.3 Results and Analysis

5.3.1 Model Performance

Based on the evaluation metrics, the Random Forest Regressor emerged as the best model with the highest R^2 score and the lowest MSE on the test set:

R^2 Score: 0.257 MSE: 0.515 The Gradient Boosting Regressor was the second-best performing model, with slightly lower R^2 and higher MSE:

R^2 Score: 0.223 MSE: 0.539

5.3.2 Comparative Performance

The following insights were derived from the comparative analysis:

Random Forest: Outperformed all other models in terms of test set performance. Demonstrated robustness against overfitting, as evidenced by consistent train/test loss.

Gradient Boosting: Exhibited high accuracy but was marginally outperformed by Random Forest. Effective at capturing non-linear relationships.

AdaBoost and Decision Tree: Showed higher test losses, indicating overfitting or poor generalization.

Linear Regression: While interpretable, it underperformed due to its inability to capture non-linear dependencies.

Support Vector Regression: Showed moderate performance but was computationally expensive.

KNN: Struggled with the complexity of the dataset, leading to suboptimal results.

5.3.3 Visualization of Results

The performance of the models was visualized using bar plots:

Mean Squared Error (MSE) Comparison:

Random Forest and Gradient Boosting achieved the lowest MSE. AdaBoost and Decision Tree had the highest MSE, indicating poorer performance.

Train vs. Test Loss:

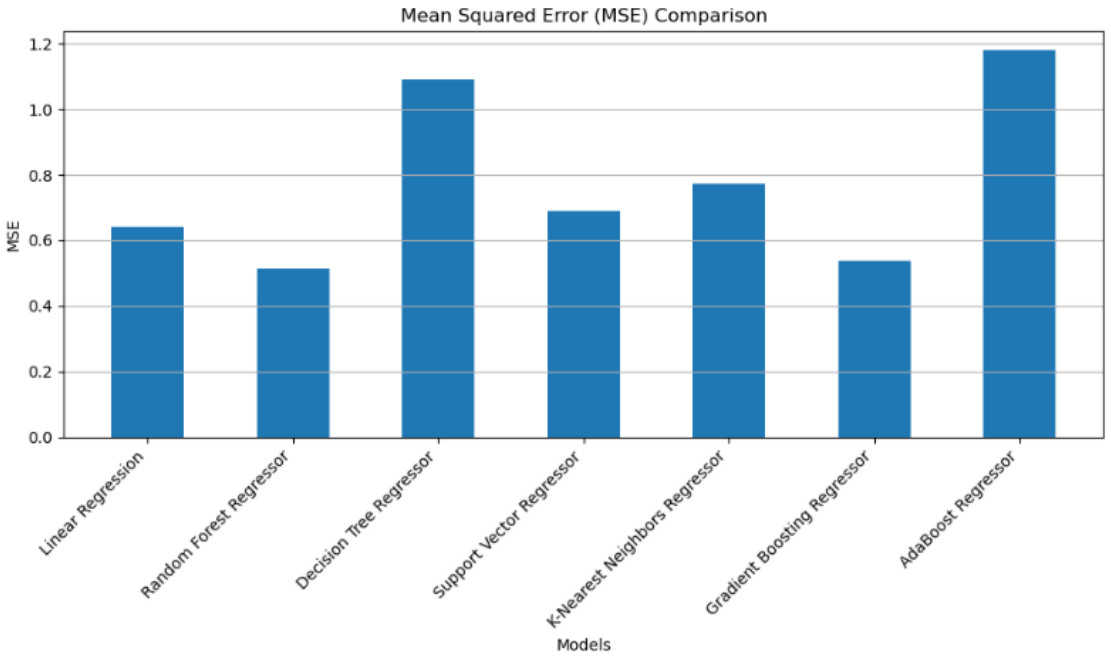

FIGURE 5.1: MSE Comparision

Random Forest and Gradient Boosting demonstrated balanced train and test losses, indicating good generalization. Models like Decision Tree showed significant discrepancies, suggesting overfitting.

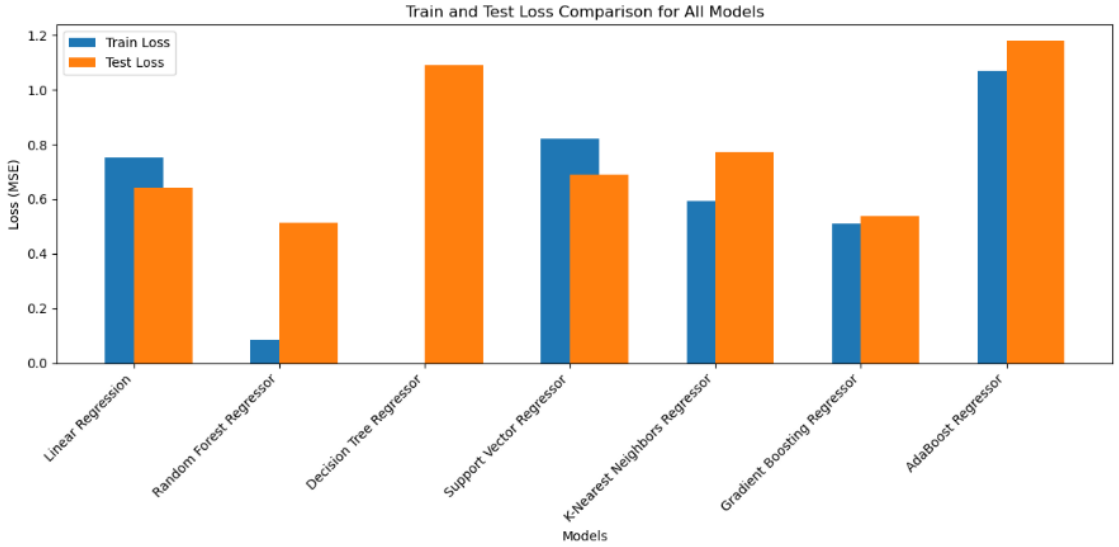

FIGURE 5.2: Train-Test MSE Loss

Discussion

The analysis highlights the following:

Ensemble methods like Random Forest and Gradient Boosting are effective for this dataset due to their ability to handle non-linear relationships and interactions between features. Simpler models like Linear Regression and KNN are less suited for complex datasets with diverse features. Train/Test Loss Comparison helped identify overfitting in models like Decision Tree and AdaBoost.

```
Best Model:
Model        Random Forest Regressor
R2_Score                     0.25715
MSE                          0.51532
Name: 0, dtype: object

Second Best Model:
Model        Gradient Boosting Regressor
R2_Score                        0.222743
MSE                             0.539189
Name: 1, dtype: object
```

FIGURE 5.3: Best Models

Summary

In this chapter, we presented and evaluated multiple machine learning models for predicting CO adsorption capacity. The Random Forest Regressor emerged as the best-performing model, balancing accuracy and generalization. The insights gained from this analysis provide a strong foundation for selecting and optimizing predictive models for MOF-based carbon capture systems

Chapter 6

Future work and Conclusion

6.1 Conclusion

The increasing need for sustainable solutions to mitigate climate change has driven significant research into carbon capture technologies. This study focused on leveraging machine learning models to predict the CO adsorption capacity of metal-organic frameworks (MOFs) based on their textural, chemical, and process properties. By compiling a comprehensive dataset and implementing advanced predictive models, this work aimed to enhance the understanding of the factors influencing carbon capture performance in MOFs.

We systematically evaluated a range of machine learning models, from simple linear regression to complex ensemble and neural network methods. The Random Forest Regressor and Gradient Boosting Regressor emerged as the most effective models, providing accurate predictions while maintaining interpretability. These models demonstrated the potential of integrating data-driven approaches with material science to streamline the discovery and optimization of MOFs for carbon capture applications.

The study also highlighted key features—such as BET surface area, pore volume, and CO partial pressure—that significantly influence adsorption capacity. However, the dataset's limitations, particularly the absence of detailed chemical properties like acidic/basic sites and functionalization, emphasize the need for more comprehensive data collection.

This work lays the groundwork for developing machine learning-driven frameworks that not only predict material performance but also provide actionable insights for designing new materials.

6.2 Future Work

While this study achieved meaningful results, several areas remain open for future exploration and improvement:

6.2.1 Incorporating Additional Features

The dataset used in this study primarily focused on textural and process-related features. Future research should aim to include additional chemical and structural properties, such as:

Acidic and Basic Sites: Quantifying the number and strength of acidic/basic sites can provide deeper insights into adsorption mechanisms.

Functionalization Details: Incorporating data on functional groups, such as amines or polar molecules, can improve the prediction of selective adsorption behavior.

Thermodynamic and Kinetic Properties: Features like enthalpy of adsorption or diffusion coefficients can enhance the predictive capability of machine learning models.

These additional features can lead to more explainable and robust models, enabling researchers to identify specific design principles for high-performance MOFs.

6.2.2 Developing Explainable Machine Learning Models

While ensemble models such as Random Forest and Gradient Boosting achieved high accuracy, their interpretability is limited. Future work should focus on integrating explainable AI techniques, such as:

SHAP (SHapley Additive exPlanations): To quantify the contribution of each feature to the prediction.

LIME (Local Interpretable Model-Agnostic Explanations): To provide local explanations for individual predictions.

These approaches will help researchers understand how specific features, such as pore diameter or functionalization, influence the adsorption performance, enabling more targeted material design.

6.2.3 Expanding the Dataset

The dataset used in this study represents an important step toward data-driven carbon capture analysis, but its scope can be expanded by:

Curating Larger Datasets: Collecting data from experimental and computational studies to improve model generalizability.

Standardizing Reporting Practices: Encouraging researchers to report adsorption data in consistent formats, including detailed information on material properties and experimental conditions.

Incorporating Multi-Gas Adsorption Data: Extending the analysis to include selectivity data for CO/N and CO/CH mixtures, which are critical for real-world applications.

Bridging Simulation and Experimentation Future work can focus on integrating experimental and computational datasets to bridge the gap between theoretical

predictions and practical applications. Combining these data sources can provide more realistic and actionable insights into MOF design and performance.

Guiding Material Design

By using machine learning models to predict and optimize material properties, researchers can accelerate the discovery of novel MOFs tailored for specific applications. Predictive models can serve as a blueprint for designing MOFs with:

Enhanced selectivity for CO over other gases. Stability under varying temperature and humidity conditions. High recyclability for cost-effective deployment. These insights will support the development of MOFs with superior performance, driving progress toward scalable and sustainable carbon capture technologies.

6.3 Closing Remarks

This study demonstrates the potential of machine learning to revolutionize material design and optimization in carbon capture technologies. By combining data-driven approaches with material science, researchers can uncover new relationships between material properties and performance, paving the way for breakthroughs in adsorption-based carbon capture. Future work, focusing on expanded datasets, enhanced feature sets, and explainable models, will further strengthen the integration of machine learning in this critical domain. Through continued innovation and collaboration, we can make significant strides toward mitigating the global climate crisis

Bibliography