Gas hydrate formation and dissociation: predictive, thermodynamic, and dynamic models

by

Mostafa Hosseini

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Earth Sciences

Waterloo, Ontario, Canada, 2025

© Mostafa Hosseini 2025

Examining committee membership

The following committee members served on the Examining Committee for this thesis.

External Examiner	Dr. Luis E. Zerpa Associate Professor, Colorado School of Mines
Supervisor	Dr. Yuri Leonenko Associate Professor, University of Waterloo
Internal Members	Dr. Andre Unger Associate Professor, University of Waterloo
	Dr. Chris Yakymchuk Associate Professor, University of Waterloo
Internal-external Member	Dr. Eric Croiset Professor, University of Waterloo

Author's declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Statement of contributions

Chapter 2

Mostafa Hosseini: Conceptualization, Methodology, Investigation, Validation, Writing – original draft. Yuri Leonenko: Supervision, Conceptualization, Methodology, Funding acquisition, Writing – Reviewing and Editing.

Hosseini, M., & Leonenko, Y. (2023). A reliable model to predict the methane-hydrate equilibrium: An updated database and machine learning approach. *Renewable and Sustainable Energy Reviews*, *173*, 113103. https://doi.org/10.1016/j.rser.2022.113103

This chapter is reproduced in full with permission of the author(s), in accordance with Elsevier's Author Rights policy.

Chapter 3

Mostafa Hosseini: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Richard Boudreault: Writing – review & editing, Project administration, Funding acquisition, Conceptualization. Yuri Leonenko: Writing – review & editing, Supervision, Investigation, Funding acquisition, Conceptualization.

Hosseini, M., Boudreault, R., & Leonenko, Y. (2025). Insights into predicting equilibrium conditions of clathrate hydrates of methane+ water-soluble hydrate former. *Fuel*, *384*, 133945. https://doi.org/10.1016/j.fuel.2024.133945

This chapter is reproduced in full with permission of the author(s), in accordance with Elsevier's Author Rights policy.

Chapter 4

Mostafa Hosseini: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Richard Boudreault: Writing – review & editing, Project administration, Funding acquisition, Conceptualization. Yuri Leonenko: Writing – review & editing, Supervision, Investigation, Funding acquisition, Conceptualization.

This chapter includes a manuscript currently under submission to *Energy & Fuels Journal* (ACS Publications). Chapter 4 is based on a manuscript submitted to Energy & Fuels (American Chemical Society), and is included here to fulfill the degree requirements of the University of Waterloo. This version of the manuscript has not been accepted for publication at the time of thesis submission. If the article is published, the final version will be available via the ACS Publications website. Reprinted with permission from the author(s). Copyright © 2025 by the authors.

Chapter 5

Mostafa Hosseini: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Richard Boudreault: Writing – review & editing, Project administration, Funding acquisition, Conceptualization. Yuri Leonenko: Writing – review & editing, Supervision, Investigation, Funding acquisition, Conceptualization.

This chapter has been prepared for publication.

Abstract

Gas hydrates are a type of crystalline compound consisting of water and small gas molecules. A wide range of applications of gas hydrates in storing natural gas in the form of artificially created solid hydrates, known as solidified natural gas technology, gas separation processes, and seawater desalination technology, has attracted great interest in scientific and practical studies. Gas hydrate formation may also cause deleterious effects, such as blockage of gas pipelines. Therefore, accurate prediction of equilibrium conditions for gas hydrates is of great interest. In this regard, machine learning-based models were proposed to predict methane-hydrate formation temperature for a wide range of brines. A comprehensive database including 987 data samples covering 15 different brines was gathered from the literature. After data cleaning and preparation, three different models, namely multilayer perceptron (MLP), decision tree (DT), and extremely randomized trees (ET), were trained and tested. The ET model achieved the best performance with a root mean squared error (RMSE) of 0.6248 K for the testing dataset. Moreover, in an additional independent testing with $MgBr_2$ samples, ET achieved an RMSE of 0.3520 K, confirming its strong generalization ability. The order of model accuracy was ET greater than MLP greater than DT. Compared to previous studies, the developed models achieved similar or better accuracy while covering a wider range of brine types. The findings of this study can be used as a reliable tool to predict methane-hydrate formation PT curves for pure water, single-salt brines, and multi-salt brines.

The research further focuses on improving the prediction of equilibrium conditions in methane hydrate systems by incorporating diverse water-soluble hydrate formers and applying advanced machine learning techniques. Methane hydrates, which naturally form under high pressure and low temperature, can be more efficiently formed or dissociated by altering thermodynamic conditions using these hydrate formers. Accurate prediction of these conditions is crucial for optimizing gas storage and energy applications. Molecular descriptors and operational parameters, such as mole fraction and pressure, were used as input variables to predict equilibrium temperature. Machine learning methods, including Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Multi-Layer Perceptron (MLP), were employed, using a novel former-based data-splitting approach rather than traditional sample-based methods. The RF model achieved the best

results, with R^2 = 0.930, RMSE = 1.71, and AARD = 0.48%. Feature selection, preprocessing, and Shapley Additive Explanations (SHAP) provided valuable insights into variable importance. Additional findings from the reduced model revealed that even less influential features significantly impacted distance-based models such as SVM and MLP. Interaction analysis through SHAP dependency plots highlighted the critical interplay between polar surface area and rotatable bonds in hydrate formation conditions. This work advances methane hydrate research by offering a more accurate and interpretable framework for predicting hydrate equilibrium, addressing key gaps in previous studies, and extending its applicability to a broader range of systems. Moreover, the introduction of a former-based data-splitting method improves generalization across different hydrate formers, while the use of SHAP values for model interpretability offers deeper insights into the relationships between molecular descriptors and hydrate equilibrium conditions. This study paves the way for improved selection of hydrate formers in hydrate systems.

In addition to the phase equilibrium studies, this research also addresses the behavior of gas hydrates under confinement, focusing on hydrate dissociation in porous media. Understanding the dissociation behavior of gas hydrates in confined porous media is crucial for evaluating their stability and potential applications in energy storage, carbon capture, and climate modeling. Two distinct approaches were developed, namely a thermodynamic activity model and machine learning (ML) models, to predict equilibrium dissociation temperatures of gas hydrates in porous media of varying pore sizes. The activity model accounted for capillary effects and surface interactions and was validated against an unfiltered experimental dataset. For CH_4 hydrates, the model achieved an AAD% of 0.17%, and for C_3H_8 hydrates, an AAD% of 0.62%. Complementary machine learning models (DT, RF, SVM, MLP) were trained using pore diameter, pressure, and gas critical properties as features. Group-based data splitting, with propane data reserved for testing, ensured robust evaluation. Among ML models, the SVM achieved the best predictive performance with an AAD% of 0.52%. SHAP analysis revealed that critical temperature, system pressure, and pore size were dominant predictors. The study also noted that experimental scatter was linked to pore structure variability and procedural differences, with larger pores showing convergence to bulk hydrate behavior. The combined modeling framework effectively captures hydrate behavior across a wide range of confined conditions, offering valuable predictive capability for both industrial and geological hydrate systems. In conclusion, the integration of

physics-based and data-driven modeling enables accurate prediction of hydrate dissociation temperatures across a range of porous media. These findings support the development of predictive tools for hydrate systems in both geological and industrial applications.

Finally, to complement the thermodynamic and equilibrium predictions, the dynamic transport behavior of hydrate particles in pipelines was investigated through CFD–DEM simulations. The dynamic behavior of hydrate particles suspended in water-dominated horizontal pipe flow using a two-way coupled CFD–DEM framework based on OpenFOAM and LIGGGHTS via CFDEM® coupling was explored. Multiphase flow simulations were conducted across inlet velocities of 0.2, 0.5, and 0.8 m/s and hydrate volume fractions of 2%, 5%, 8%, 15%, and 20%. Pressure drop behavior was quantified by extracting pressure gradients between two axial positions ($z = 0.10$ m and $z = 0.49$ m) early in the simulation. Results indicated that pressure drop increases with hydrate volume fraction at all flow velocities, with clustering phenomena becoming more prominent at higher solid loadings. Cross-sectional velocity profiles visualized the early evolution of particle clustering, wall interactions, and domain depletion. Increased flow velocity enhanced particle suspension but reduced domain uniformity over time. Time-resolved analyses of pressure drop, drag force, particle velocity, interparticle forces, and radial migration were conducted to explore flow regime transitions and mechanical resistance. Early clustering near the pipe walls was observed under dense flow conditions, driven by cohesive and frictional forces, leading to partial stratification and localized energy dissipation. The study highlights the importance of considering early-time flow dynamics, where suspension quality and transport resistance are most sensitive to hydrate loading. These findings contribute to a deeper understanding of hydrate slurry transport in multiphase pipeline systems and offer practical guidance for improving flow assurance models and mitigation strategies in subsea energy operations.

Acknowledgements

I am deeply grateful to my supervisor, Dr. Yuri Leonenko, for his exceptional guidance and unwavering support throughout my academic journey. From the very beginning, his encouragement to think critically, conduct independent research, and adopt interdisciplinary approaches has been invaluable to my development as a researcher.

I would also like to sincerely thank the members of my examining committee: Dr. Andre Unger, Dr. Eric Croiset, and Dr. Chris Yakymchuk for their insightful feedback and constructive comments, which greatly enhanced the quality of this work. Special thanks to Dr. Luis E. Zerpa for serving as the external examiner and for generously sharing his expertise, further enriching this research.

Financial support for this research was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), Mitacs, and AWN Nanotech's support, thanks to the contribution of Richard Boudreault, all of which are gratefully acknowledged.

Finally, I wish to express my heartfelt gratitude to my wife, my parents, my sister, and my brothers for their unconditional love, encouragement, and support. I dedicate this thesis to them, to whom I owe everything.

Table of contents

Examining committee membership ii

Author's declaration iii

Statement of contributions iv

Abstract vi

Acknowledgements ix

List of figures xiii

List of tables xiv

Chapter 1: Introduction 1

1.1. Background and significance 1

1.2. Hydrate thermodynamics and kinetics 3

1.3. Hydrate transport and flow assurance 4

1.4. Thesis objectives 5

1.5. Thesis outline 7

Chapter 2: A reliable model to predict the methane-hydrate equilibrium: an updated database and machine learning approach 10

2.1. Introduction 10

2.2. Method 12

2.3. Database development 13

2.3.1. Literature survey 13

2.3.2. Data quality check 17

2.3.3. Feature selection for modeling 18

2.3.4. Data splitting 18

2.4. Result and discussion 19

2.4.1. Model development 19

2.4.2. Accuracy assessment 21

2.4.3. Additional testing 23

2.4.4. Comparison with literature 24

2.5. Conclusions 25

2.6. Supplementary data 26

Chapter 3: Insights into predicting equilibrium conditions of clathrate hydrates of methane + water-soluble hydrate former 27

3.1. Introduction 27

3.2. Database 32

3.3. Methodology 36

3.3.1. Models 36

3.3.2. Molecular descriptors 37

3.3.3. Data splitting 40
3.3.4. Feature transformation 41
3.3.5. Feature filtering 42
3.3.6. Scaling 46
3.3.7. SHapley Additive exPlanations (SHAP) method 47
3.4. Results and discussion 49
3.4.1. Model development 49
3.4.2. Accuracy assessment 51
3.4.3. Reduced model 53
3.4.4. Model complexity reduction by feature selection 56
3.5. Conclusions 60
Chapter 4: Gas hydrate dissociation temperature prediction in porous media: effects of pore size and modeling approach 63
4.1. Introduction 63
4.2. Models 67
4.2.1. Activity model 67
4.2.2. Predictive models 71
4.3. Result and discussion 74
4.3.1. Activity model 77
4.3.2. Predictive models 82
4.4. Conclusions 86
Chapter 5: Investigation of hydrate particle dynamics and distribution in water-dominated flows 88
5.1. Introduction 88
5.2. Methodology 92
5.2.1. Fluid phase modeling 93
5.2.2. Particle motion modeling 94
5.2.3. Fluid–particle interaction forces 94
5.2.4. Simulation setup and assumptions 95
5.2.5. Coupling strategy 97
5.3. Results and discussion 97
5.3.1. Pressure drop response to hydrate loading and flow rate in horizontal pipe flow 98
5.3.2. Particle retention and depletion 101
5.3.3. Pressure evolution 102
5.3.4. Fluid velocity trends 104
5.3.5. Drag force behavior 105
5.3.6. Interparticle force dynamics 107
5.3.7. Particle velocity trends 109

5.3.8. Radial distribution of particles 111
5.3.9. Statistical overview of particle-phase behavior 113
5.3.10. Velocity-based visualization of particle distribution: temporal and concentration effects 114
5.4. Conclusions 118
Chapter 6: Conclusions and recommendations 119
6.1. Summary of key findings 119
6.2. Recommendations for future research 121
References 124

List of figures

Figure 1-1: Three different gas hydrate structures and their corresponding cages [3].....2
Figure 1-2: Pressure-temperature diagram for hydrate phases: regions of hydrate formation and dissociation [16].....3
Figure 2-1: The P-T diagram for (a) NaBr and (b) NaCl+$CaCl_2$ systems.....16
Figure 2-2: The box-plot for the 5-fold CV RMSE acquired from 20 trials for MLP, DT, and ET models. 20
Figure 2-3: comparison of predicted and experimental T for the testing dataset. (a) cross-plot and (b) Residual plot. 23
Figure 2-4: Comparison among different models for predicting equilibrium T for $MgBr_2$ system (a) cross-plot and (b) residual plot.....24
Figure 3-1: 2D molecular structures of each former included in the gathered database. 35
Figure 3-2: Correlation matrix of selected features. 45
Figure 3-3: Flowchart of research methodology. 48
Figure 3-4: The violin plot for the results acquired from 100 trials for the ML models (a) 5-fold CV RMSE and (b) fitting time. 51
Figure 3-5: Feature importance metrics based on the developed RF model. 53
Figure 3-6: Systematic comparison of the original and reduced models based on (a) AARD (%), (b) RMSE, and (c) MAE. 55
Figure 3-7: Beeswarm plot visualizing the distribution of SHAP values for key predictive features. 57
Figure 3-8: SHAP value dependency plots for features affecting hydrate equilibrium temperature prediction. 59
Figure 3-9: (a) Interaction heatmap matrix for variables, highlighting strength of effects. (b) SHAP values for psa4 colored by numRT0 levels. 60
Figure 4-1: Flowchart of calculating hydrate dissociation pressure in a porous medium using the activity model. 71
Figure 4-2: Overview of the experimental data used in this study showing equilibrium dissociation conditions of (a) CH_4 [32–37] and (b) C_3H_8 [37,38] hydrates across a range of porous media with different pore sizes. 76
Figure 4-3: Experimental data and predicted CH_4 hydrate dissociation temperatures in porous media with different pore diameters: (a) 6 nm, (b) 10 nm, (c) 15 nm, (d) 30 nm, (e) 50 nm, and (f) 100 nm, using the activity model. 77
Figure 4-4: SHAP beeswarm plot showing feature impacts on SVM predictions. 84
Figure 4-5: Scatter plots of input features vs. predicted equilibrium temperature. 85
Figure 5-1: Schematic of the computational domain with hydrate particle distribution. 95
Figure 5-2: Pressure drop vs. volume fraction in horizontal hydrate–water flow. 100
Figure 5-3: Number of particles vs. Time for different hydrate volume fractions at inlet velocities of (a) 0.2 m/s, (b) 0.5 m/s, and (c) 0.8 m/s. 101
Figure 5-4: Figure 3. Average pressure vs. time for hydrate–water flow at five volume fractions ((a) 2%, (b) 5%, (c) 8%, (d) 15%, and (e) 20%) with three inlet velocities (0.2, 0.5, and 0.8 m/s). 103
Figure 5-5: Average axial velocity (z-direction) vs. time for hydrate–water flow at fixed volume fractions (a) 2%, b) 5%, c) 8%, d) 15%, and e) 20%) with varying inlet velocities (0.2, 0.5, and 0.8 m/s). 105
Figure 5-6: Average fluid–particle drag force vs. time for hydrate–water flow at fixed volume fractions (a) 2%, b) 5%, c) 8%, d) 15%, and e) 20%) and varying inlet velocities (0.2, 0.5, and 0.8 m/s). 106
Figure 5-7: Average interparticle contact force vs. time at fixed hydrate volume fractions (a) 2%, b) 5%, c) 8%, d) 15%, and e) 20%), evaluated for three inlet velocities (0.2, 0.5, and 0.8 m/s). 108
Figure 5-8: Average particle velocity vs. time for hydrate–water flow at fixed volume fractions (a) 2%, b) 5%, c) 8%, d) 15%, and e) 20%) and inlet velocities of 0.2, 0.5, and 0.8 m/s. 110
Figure 5-9: Average radial distance from the pipe center for hydrate particles over time, shown for five volume fractions (a) 2%, b) 5%, c) 8%, d) 15%, and e) 20%) at three inlet velocities (0.2, 0.5, and 0.8 m/s).....112
Figure 5-10: 2D cross-sectional velocity plots at the pipe inlet (z = 0 m) for hydrate–water flow at a volume fraction of 15% and velocity of 0.5 m/s. Snapshots are shown at time intervals of a) 0.08 s, b) 0.24 s, c) 0.56 s, d) 0.72 s, e) 0.88 s, and f) 0.96 s.....115
Figure 5-11: Cross-sectional inlet views of hydrate particle velocity magnitude at a fixed simulation time of 0.24 s and a flow velocity of 0.5 m/s. Volume fractions considered are a) 2%, b) 5%, c) 8%, d) 15%, and e) 20%. 117

List of tables

Table 2-1: The sources of the gathered database with the number of data samples for each brine. 13
Table 2-2: The ranges for brine concentrations based on both salt and ion types. 15
Table 2-3: The present duplicates in the gathered database. 17
Table 2-4: The determined hyperparameters as well as the searching space for each model. 20
Table 2-5: The statistical parameters denoting the accuracy of the developed models in this study. 22
Table 2-6: Statistical error analysis of different proposed models for the $MgBr_2$ system. 24
Table 2-7: Comparison between the models developed in this study and literature models regarding the RMSE for the testing dataset. 25
Table 3-1: Database of chemical compounds: physicochemical properties and experimental conditions. 33
Table 3-2: Molecular descriptors used in this study, employing the RDKit library [156]. 38
Table 3-3: Statistical parameters of the molecular descriptors used in this study. 40
Table 3-4: Transformations used in this study. 41
Table 3-5: The skewness and kurtosis for the original and transformed formers' descriptors. 43
Table 3-6: The skewness and kurtosis for the original and transformed operational parameters. 46
Table 3-7: The determined hyperparameters as well as the searching space for each model. 50
Table 3-8: The statistical parameters showing the accuracy of the models developed in this study. 52
Table 3-9: The determined hyperparameters as well as the searching space for each reduced model. 54
Table 3-10: Comparison of the testing RMSE between the original and reduced models. 55
Table 4-1: Langmuir Constant Parameters for CH_4 and C_3H_8 in Equation 4-4 [66]. 68
Table 4-2: Thermodynamic characteristics at reference conditions for gas hydrates [2]. 70
Table 4-3: Henry's law constants [195]. 71
Table 4-4: Summary of the data used for the training and testing sets, including gas types, pressure and temperature ranges, pore diameters, and sample counts. 73
Table 4-5: The calculated AAD for predicting the dissociation temperatures of CH_4 and C_3H_8 hydrates in porous media using the activity model. 78
Table 4-6: Optimal hyperparameters and performance metrics from randomized search. 83
Table 4-7: Metrics employed to measure the accuracy of the models developed in this work. 83
Table 5-1: Fluid properties. 97
Table 5-2: Particle properties. 97
Table 5-3: Averaged pressure drop and pressure gradient ($\Delta P/L$) for hydrate–water mixtures at varying flow rates and volume fractions. 98
Table 5-4: Truncated (0–0.6 s) time-averaged metrics for hydrate–water flows at varying flow rates and volume fractions. 113

Chapter 1: Introduction

1.1. Background and significance

Clathrate hydrates are non-stoichiometric crystalline compounds composed of small gas molecules such as methane, ethane, carbon dioxide, and xenon encapsulated within hydrogen-bonded water cages [1]. These structures typically form under high-pressure and low-temperature conditions, just above the freezing point of water. Interest in gas hydrates spans two major domains: their problematic formation during natural gas transportation, where they can cause pipeline blockages, and their potential applications in energy storage, carbon capture, seawater desalination, and gas separation technologies [1,2]. Despite their promise, several challenges hinder the large-scale application of hydrate-based technologies, including the difficulty of controlling nucleation and optimizing formation conditions under moderate operating environments.

Gas hydrates crystallize into three predominant structures: Structure I (sI), Structure II (sII), and Structure H (sH) (Figure 1-1) [1]. The specific structure formed depends largely on the van der Waals radii of the guest molecules. Structure I is typically stabilized by smaller molecules such as methane and carbon dioxide, comprising two small (5^{12}) and six large ($5^{12}6^2$) cages per unit cell. Structure II forms with slightly larger molecules like propane, containing 16 small (5^{12}) and 8 large ($5^{12}6^4$) cages per unit cell. Structure H, the most complex, requires a combination of large and small guest molecules and features three distinct types of cages [1,2]. Typically, gas hydrates contain approximately 85% water by mole, giving them physical properties similar to those of ice [2].

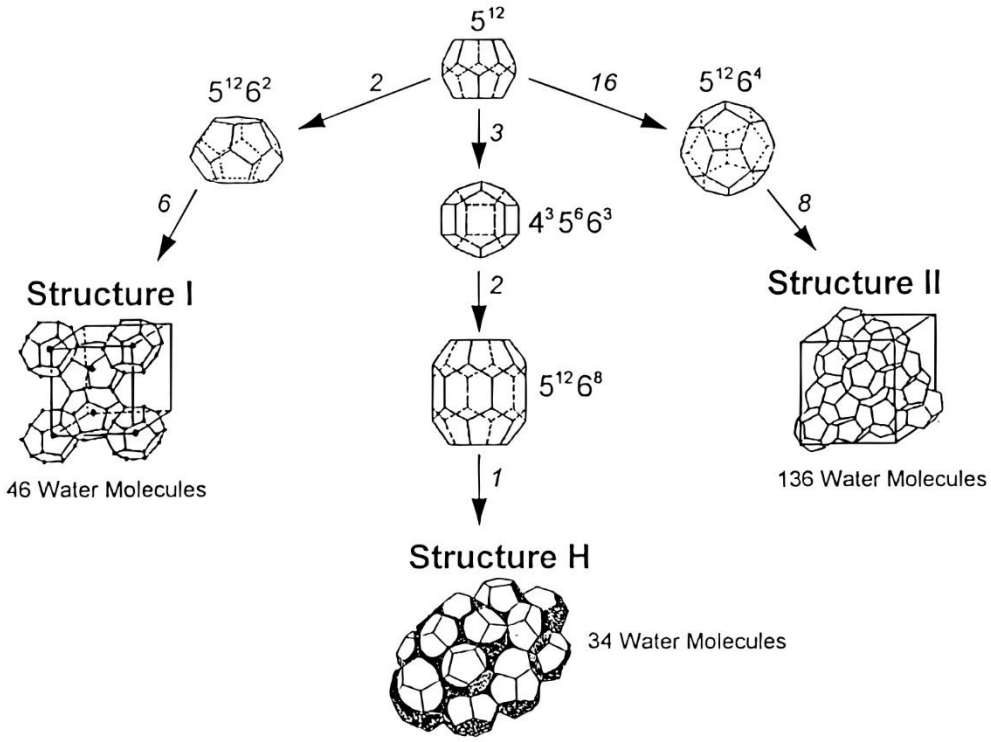


Figure 1-1: Three different gas hydrate structures and their corresponding cages [3].

Reproduced with permission from A.A. Khokhar, J.S. Gudmundsson, E.D. Sloan, "Gas storage in structure H hydrates," Fluid Phase Equilibria, Vol. 150, 1998, Figure 1. © Elsevier.

Although hydrate formation presents operational challenges for the oil and gas industry [4], the unique properties of gas hydrates offer significant technological opportunities. Due to their high gas-to-solid volume ratio (approximately 170 v/v) [2], gas hydrates are considered promising candidates for natural gas storage and transportation in the form of Solidified Natural Gas (SNG) [2]. Furthermore, the self-preservation phenomenon, where a protective ice layer forms at around 253 K, enables stable storage at near-atmospheric pressures [5]. In addition to storage applications, gas hydrates are being explored for gas separation [6,7], carbon dioxide capture [8,9], secondary refrigeration [10,11], and seawater desalination [12,13].

1.2. Hydrate thermodynamics and kinetics

A thorough understanding of both the thermodynamics and kinetics of gas hydrates is essential for advancing hydrate technologies. Hydrate formation involves two main stages: nucleation and growth, while dissociation is typically triggered by heat input or pressure reduction [14,15]. These kinetic factors directly influence the rates of hydrate formation and dissociation under different environmental and operational conditions. Therefore, integrating both thermodynamic and kinetic considerations into predictive models is critical for developing reliable flow assurance strategies.

The phase behavior of gas hydrates is conventionally represented by pressure-temperature equilibrium diagrams, which delineate the stability regions for hydrate formation and dissociation (Figure 1-2) [16]. The presence of salts and other impurities can significantly affect hydrate stability by reducing water activity and shifting equilibrium conditions [17]. Thermodynamic Hydrate Inhibitors (THIs), such as methanol and glycols, suppress hydrate formation by lowering the water activity, whereas Thermodynamic Hydrate Promoters (THPs), such as tetrahydrofuran (THF) and cyclopentane, facilitate hydrate formation by altering the chemical potential of water [17,18].

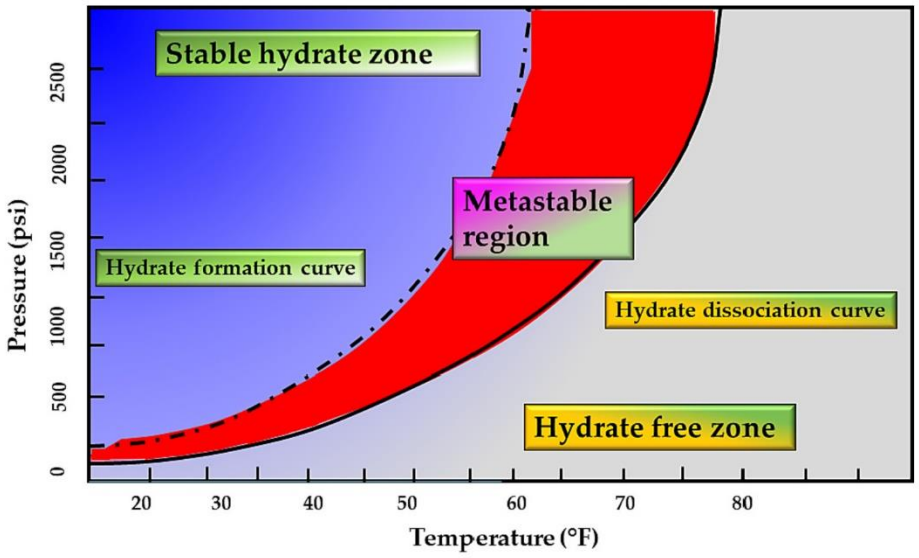


Figure 1-2: Pressure-temperature diagram for hydrate phases: regions of hydrate formation and dissociation [16].

Reproduced from Abbasi et al., Earth-Science Reviews, 2022 [16], licensed under *CC BY 4.0.*

Accurate prediction of hydrate phase equilibria is important for both fundamental research and industrial applications. The van der Waals–Platteeuw (vdWP) model [19–24], together with modifications such as the Kihara potential correction [25], has significantly improved the understanding of hydrate stability. However, the performance of these models decreases in systems involving complex gas mixtures, inhibitors, or promoters [26]. In recent years, machine learning methods have emerged as powerful alternatives, showing superior predictive capabilities for hydrate formation and dissociation conditions in complex and multicomponent systems [27–31].

In porous media, hydrate formation and stability are affected by confinement effects. Factors such as capillary pressure, interfacial tension, and surface interactions shift the equilibrium conditions to higher pressures compared to bulk systems [32–38]. Experimental studies using silica gels with different pore sizes have confirmed these effects [38]. Corresponding modeling approaches have incorporated parameters such as capillary effects, activity coefficients, and pore size distributions to predict hydrate stability in confined environments more accurately [39].

Despite substantial progress in thermodynamic modeling of hydrate systems, several important limitations remain. Traditional models often perform poorly when applied to complex systems involving mixed salts, multiple hydrate formers, or hydrates formed under confinement conditions. They typically neglect the influence of individual parameters, and the specific effects of each contributing factor should be systematically identified and explicitly considered. In addition, existing models are frequently developed for specific salts or limited experimental conditions, which restricts their generalizability to more diverse systems. Similarly, while machine learning methods have shown significant promise in predicting hydrate equilibria, most prior studies have focused on specific salts or gas formers, resulting in models that are only applicable to a restricted range of systems. These models often rely on limited datasets and demonstrate poor extrapolation capability. Furthermore, the lack of interpretability in many machine learning approaches presents a barrier to their wider adoption in scientific and engineering practice.

1.3. Hydrate transport and flow assurance

The transport behavior of gas hydrates in water-dominated multiphase flow systems presents significant challenges for subsea pipeline operations. Hydrate formation, growth, aggregation, and

deposition in the flowlines greatly increase the risk of blockages. This makes it essential to understand and assess hydrate behavior in these systems to ensure safe and reliable production [40,41] . Recent advances in Computational Fluid Dynamics coupled with Discrete Element Methods (CFD-DEM) have improved the understanding of hydrate particle dynamics, including their transport, aggregation, and deposition mechanisms. Key operational parameters, such as flow velocity, hydrate volume fraction, particle-particle interactions, and pipe geometry, have been shown to strongly influence slurry behavior [42,43]. Observations indicate that hydrate clustering near the pipe walls, slurry stratification, and associated increases in pressure drop serve as early indicators of potential blockages [43]. Consequently, the development of predictive CFD–DEM models has become essential for designing effective flow assurance strategies to mitigate hydrate-related risks in subsea pipelines [44]. Despite ongoing research, hydrate transport in pipelines, particularly the transient clustering, wall interactions, and phase transitions that occur under multiphase flow conditions, remains poorly understood. These phenomena are difficult to capture using conventional modeling approaches, especially during the early stages of flow development. There is a clear need for detailed, time-resolved simulations capable of accurately capturing particle–fluid and particle–wall interactions. In particular, open-source CFD–DEM tools offer a promising but not yet widely used approach to advancing the understanding of hydrate slurry transport under realistic flow conditions.

1.4. Thesis objectives

To address the knowledge gaps and modeling limitations identified in the introduction, this study is structured around a set of targeted objectives that directly respond to those challenges. Robust thermodynamic and data-driven models were developed, and physical interpretability tools such as SHapley Additive Explanations (SHAP) were employed to improve understanding of the data-driven predictions. In addition, time-resolved simulations of hydrate–water flows were conducted to overcome limitations in generalizability, early-time flow characterization, and coupled particle–fluid interaction modeling. The primary aim is to enhance predictive accuracy, physical insight, and practical relevance in the context of hydrate phase behavior and transport.

The specific objectives and key innovations achieved in this work are outlined below.

1) Develop accurate machine learning models to predict methane hydrate formation conditions across pure water, single-salt, and multi-salt brine systems.

   A comprehensive and curated experimental database was compiled, consisting of 987 data points covering 15 different brines, representing the largest and most diverse dataset currently available for methane hydrate equilibrium. Machine learning models, including Multilayer Perceptron (MLP), Decision Tree (DT), and Extremely Randomized Trees (ET), were trained and validated for high predictive accuracy. The key innovation lies in integrating diverse saline systems into a unified modeling framework that is not limited to specific salts. This significantly enhances the generalizability and practical utility of the models across a wide range of aqueous environments relevant to hydrate formation.

2) Enhance the prediction of methane hydrate equilibrium conditions in the presence of water-soluble hydrate formers using advanced machine learning techniques.

   Molecular descriptors, mole fraction, and pressure were used as input features for the development of ML models such as Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), and Multilayer Perceptron (MLP). A novel former-based data-splitting strategy was introduced to enhance model generalization across different chemical formers, addressing a major limitation of traditional sample-based splitting methods. This approach improved the predictive performance and broadened the applicability of ML models for hydrate systems involving a wide range of hydrate formers.

3) Predict gas hydrate dissociation temperatures in porous media using thermodynamic and machine learning models.

   A thermodynamic activity model was developed to account for capillary pressure and pore confinement effects, validated against experimental dissociation data. Independently, machine learning models were trained using pore size, system pressure, and critical gas properties as input features, providing predictive tools for hydrate dissociation behavior in porous media.

4) Analyze hydrate particle dynamics and flow behavior in water-dominated multiphase pipeline systems using CFD–DEM simulations.

A detailed numerical framework was developed using a two-way coupled Computational Fluid Dynamics–Discrete Element Method (CFD–DEM) approach. Time-resolved simulations revealed how variations in hydrate volume fraction and inlet velocity affect particle clustering, mechanical resistance, and flow regime transitions. The findings contribute to a deeper understanding of hydrate blockage mechanisms and support flow assurance strategies for subsea pipeline operations.

5) Integrate physics-based and data-driven modeling approaches to develop comprehensive and practical predictive frameworks for hydrate systems.
   The use of SHapley Additive Explanations (SHAP) for machine learning models enhances the interpretability of predictions, identifying key physical variables such as critical temperature, system pressure, and hydrate former properties, thus facilitating the application of these models to real-world hydrate management problems.

1.5. Thesis outline

This thesis is organized into five additional chapters, each addressing a key research area within the field of gas hydrates.

Chapter 2 presents a manuscript published in *Renewable and Sustainable Energy Reviews*, titled *"A Reliable Model to Predict Methane-Hydrate Equilibrium: An Updated Database and Machine Learning Approach."* This chapter summarizes the development of machine learning models for predicting methane hydrate formation temperatures across a wide range of brine compositions. A comprehensive database of 987 data points covering 15 different brines was compiled, cleaned, and analyzed. Three machine learning algorithms, namely Multilayer Perceptron (MLP), Decision Tree (DT), and Extremely Randomized Trees (ET), were trained and evaluated. The ET model demonstrated the best performance, achieving a root mean squared error (RMSE) of 0.6248. This model provides a reliable tool for predicting methane hydrate equilibrium conditions in pure water, single-salt brines, and multi-salt systems.

Chapter 3 includes a manuscript published in *Fuel*, titled *"Insights into Predicting Equilibrium Conditions of Clathrate Hydrates of Methane with Water-Soluble Hydrate Formers."* This study

focuses on improving the prediction of methane hydrate equilibrium conditions by incorporating water-soluble hydrate formers and applying advanced machine learning techniques. Molecular descriptors, mole fractions, and pressure were used as input variables. Machine learning models, including Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Multilayer Perceptron (MLP), were developed using a novel former-based data-splitting approach. The Random Forest model achieved the highest accuracy, with a coefficient of determination (R^2) of 0.930 and an RMSE of 1.71. Feature selection and SHapley Additive exPlanations (SHAP) analysis provided interpretability, offering insights into the influence of individual features on model predictions and improving generalization across different hydrate formers.

Chapter 4 presents a manuscript currently under submission to the *Energy & Fuels Journal* (ACS Publications), titled *"Hydrate Dissociation Temperature Prediction in Porous Media: Effects of Pore Size and Modeling Approach."* This chapter investigates hydrate dissociation behavior in confined porous media using both thermodynamic modeling and machine learning methods. The thermodynamic model, based on an activity coefficient approach, demonstrated excellent agreement with experimental data for methane and propane hydrates, yielding average absolute deviations (AAD%) of 0.17% and 0.62%, respectively. Complementary machine learning models were trained using pore diameter, pressure, and critical properties as input features. Among these, the Support Vector Machine model provided the highest predictive performance, achieving an AAD% of 0.52%. SHAP analysis confirmed the model's interpretability by identifying critical temperature, pressure, and pore size as the dominant factors influencing hydrate dissociation.

Chapter 5 includes a manuscript prepared for publication, titled *"Investigation of Hydrate Particle Dynamics and Distribution in Water-Dominated Flows."* This study explores the behavior of hydrate particles suspended in water-dominated horizontal pipe flow using a two-way coupled Computational Fluid Dynamics and Discrete Element Method (CFD–DEM) framework. Simulations were conducted across a range of inlet velocities and hydrate volume fractions to assess the impact of flow energy and solid loading on particle dynamics. Results showed that increased volume fraction enhanced momentum exchange, mechanical resistance, and particle clustering, while higher flow velocities promoted particle suspension. The study provides valuable insights into flow regime transitions, mechanical resistance, and hydrate slurry behavior,

contributing to the development of improved flow assurance strategies for subsea pipeline operations.

Chapter 6 concludes the thesis by summarizing the key findings, providing recommendations, and identifying potential directions for future research in the field of gas hydrate systems.

Chapter 2: A reliable model to predict the methane-hydrate equilibrium: an updated database and machine learning approach

2.1. Introduction

Gas hydrates are a type of crystalline compounds that consists of water and small gas molecules known as the host and guest molecules, respectively. These compounds are formed at low temperatures and high pressures [2]. CH_4 (methane), C_2H_6 (ethane), N_2 (nitrogen), CO_2 (carbon dioxide), H_2 (hydrogen), and H_2S (hydrogen sulfide) can be mentioned as guest molecules in Gas hydrates [45]. The formation of gas hydrates occurs when guest molecules are trapped inside the cage structures created by host molecules [2]. Crystal structures of gas hydrates can be classified into three classes sI, sII, and sH, each of which is composed of different polyhedral combinations, where each small gas molecule affects the gas hydrate equilibrium and structure [2,45,46].

Natural gas hydrates can be mentioned as one of the energy sources because they have been discovered in many offshore and permafrost regions [47,48]. Natural gas is the cleanest fossil fuel which plays a crucial role in supplying the world's energy needs. Natural gas contains approximately 90% and in most cases much more methane. Among fossil fuels, the use of natural gas as the primary energy will continue to grow until 2040. High calorific value, low carbon emissions, and the improvement of electricity generation efficiency, compared to coal and gasoline, increases the use of natural gas, especially for electricity generation applications. More clearly, the use of natural gas for electricity generation reduces CO2 emissions by about 50% and 33%, respectively, compared to the use of coal and oil [49]. Thus, efficient options for natural gas storage and transportation are critical around the world [50]. The benefits that come with storing natural gas in the form of artificially created solid hydrates, known as solidified natural gas (SNG), provide a unique opportunity to store and ship it around the world. The SNG technology provide us with various advantages such as high volume storage capacity, environmental friendliness, high safety, cost-effectiveness, and large-scale natural gas storage [51].

It is noteworthy to mention that the SNG technology focuses on the use of methane gas due to the availability of this gas. Also, a significant portion of research over the past three decades has focused on methane (natural gas) storage in hydrates [49]. Moreover, as salts have a significant

effect on both the thermodynamics and kinetics of methane hydrates, much attention has recently been directed to the use of seawater in addition to pure water for the formation of methane hydrates [51]. In addition, the use of seawater can improve the methane hydrates formation economics [52].

Other applications of gas hydrates include their use in gas separation processes using selective partitioning of the target gas between the hydrate crystal phase and gas phase [6,7,53]. On the other hand, the formation of natural gas hydrates in an offshore wellbore and in oil and gas pipelines within cold regions causes flow-assurance problems [54,55]. Therefore, accurate prediction of equilibrium conditions for gas hydrates is necessary for a variety of reasons. Thus far, temperatures and pressures of gas hydrate equilibrium have been obtained from experiments [2], thermodynamics models [19–23], empirical correlations [56–63], and artificial intelligence techniques [27,28].

In some cases, experimental measurements of gas hydrate equilibrium conditions are costly and time-consuming. The van der Waals and Platteeuw (vdW-P) model, which is based on classical statistical thermodynamics, is one of the most common thermodynamic models for predicting the equilibrium conditions of gas hydrates [64]. However, the vdW-P model also has its limitations [64,65]. For instance, calculating the Langmuir adsorption constants is complex [66,67]. In addition, the vdW-P model has a high error in calculating the gas-hydrate equilibrium at low and high temperatures [68]. Apart from the vdW-P model, other thermodynamic models based on the equations of state have been developed [69,70]. However, estimating the intermolecular interaction parameters of equations of state is also complex [71]. In addition, these thermodynamic models do not have acceptable predictions at low temperatures and high pressures [72]. Although empirical correlations are regularly updated and are straightforward to use, employment of them for mixed gas hydrates over a wide range of temperatures and pressures leads to considerable errors and thus limits their applications in industrial problems [63,64,68].

In recent years, a number of researchers have examined the application of machine learning in the development of prediction models because these techniques have shown promising prediction accuracy [73–76]. Recently, the application of machine learning techniques in predicting gas hydrate formation conditions has received more attention. Soroush et al. (2015) proposed a model based on the artificial neural network (ANN) method to estimate the hydrate formation temperature for different gas mixtures. They utilized 279 experimental data points for construction of their

network. They were able to show that the predictions of the proposed ANN model were reasonably consistent with the experimental data [27]. Mesbah et al. (2017) developed a model based on the (least square support vector machine) LSSVM algorithm to predict hydrate formation temperature for a wide range of natural gas mixtures using 279 experimental data points. Their model presented a high accuracy with a squared correlation coefficient (R^2) of 0.9918 [77]. In the following, Amar (2021) modeled the hydrate formation temperature of natural gases using gene expression programming (GEP) model. A database containing 279 experimental data points was employed. The developed GEP model could accurately estimate the hydrate formation temperature an average absolute relative error (AARE) of 0.1397% [78]. The employed database was similar to that of Mesbah et al. (2017) [77]. More recently, Xu et al. [29] used five different machine learning predictive tools, including multiple linear regression, support vector regression, k -nearest neighbor, gradient boosting regression, and random forest, for predicting CH_4-hydrate formation condition in different brines. They used 702 data samples in this regard and achieve excellent results. Among their models, gradient boosting regression performed better than the other models.

The purpose of this study is to propose machine learning based models to predict CH_4-hydrate formation temperature for a wide range of brines. In this regard, firstly, a comprehensive database was gathered from the open literature. The database includes 987 data samples for pure water, 15 different single-salt brines, and 6 different multi-salt systems. Three different predictive models of Multilayer Perceptron (MLP) ANN, Decision Tree (DT), and Extremely Randomized Trees (ET) were implemented to accurately predict the methane-hydrate formation PT data. The performance of the developed models was also evaluated by comparison with the literature.

2.2. Method

In this study, three different predictive models of Multilayer Perceptron (MLP) ANN, Decision Tree (DT), and Extremely Randomized Trees (ET) are considered. The Scikit-Learn library [79] developed for Python is used for model implementation. For the MLP, the function sklearn.neural_network.MLPRegressor() is used. MLP is a black-box model and the fitting process is not explicitly shown. The structure of MLP includes one input layer, one or more hidden layer(s), and one output layer. The number of neurons in the input and output layers corresponds to the number of input and output features. The number of neurons in the hidden layer(s) is arbitrary. For

the DT, the function of sklearn.tree.DecisionTreeRegressor() is used. This predictive tool is of non-parametric models. DT works through developing simple decision rules based on the input variables. As DT can be visualized, it can be considered as a with-box model. However, if the DT is too deep, it will lose its simplicity to be visually understood and interpreted. For the ET, the function of sklearn.ensemble.ExtraTreesRegressor() is used. ET is an ensemble method that benefits from averaging several DTs to provide a superior result and minimize the over-fitting probability. More details about the used predictive methods are available in the literature [79,80].

2.3. Database development

2.3.1. Literature survey

A comprehensive literature review was performed to gather all the data available on the CH_4-hydrate phase equilibria. In this regard, different studies, including those that had a database [29,58,81–103] were recognized containing relevant data. The gathered database includes 987 data samples for pure water, 15 different single-salt brines, and 6 different multi-salt systems. Covering 15 different salts, it is the largest database gathered on the topic of interest so far. A comprehensive introduction to the gathered data is provided. The details of the gathered data are elaborated in Table 2-1. This table lists the publications and corresponding data. As one can see, NaCl has been the most studied brine in the literature as 12 different studies reported data on CH_4-hydrate phase equilibria for this system. Table 2-2 displays the range of salt concentrations in brines expressed as weight percentages of salt as well as the range of mole fractions of ions in salts for the database compiled from the literature. Among the studied brines, KCl has the broadest salt concentration (C) in the range of 2.6-40 %wt.

Table 2-1: The sources of the gathered database with the number of data samples for each brine.

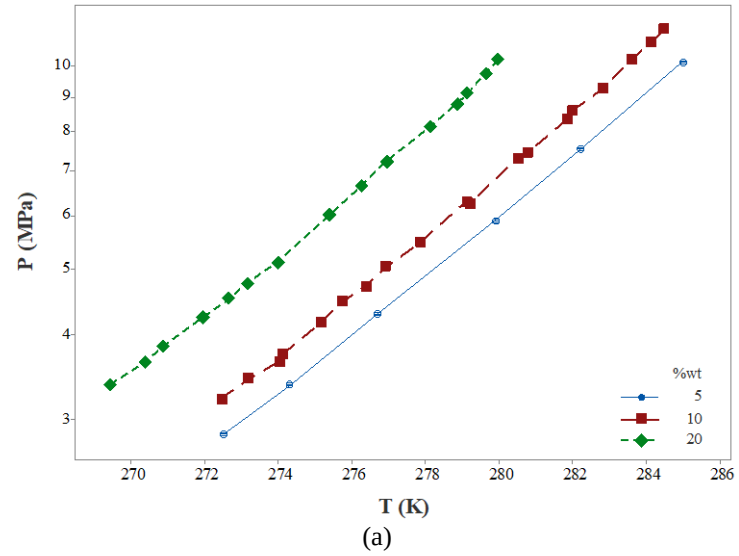
Index	Brine	Index	Year	Reference	N
		Pure Water			
1	–	1	1983	de Roo et al. [81]	9
		2	1991	Adisasmito et al. [99]	11
		3	1995	Maekawa et al. [94]	3
		4	2001	Jager and Sloan [58]	12
		5	2006	Maekawa and Imai [93]	28
		6	2017	Hu et al. [97]	6

Index	Brine	Index	Year	Reference	N
		7	2017	Kamari et al. [88]	17
		8	2021	Bhawangirkar and Sangwai [101]	13
		Single Salt			
1	$CaBr_2$	1	2009	Mohammadi et al. [96]	10
		2	2017	Hu et al. [86]	6
		3	2021	Bhawangirkar and Sangwai [101]	41
2	$CaCl_2$	1	2003	Kharrat and Dalmazzone [90]	26
		2	2006	Atik et al. [100]	6
		3	2008	Mohammadi et al. [95]	12
		4	2018	Hu et al. [87]	20
		5	2019	Aregbe et al. [98]	21
		6	2019	Du et al. [83]	14
		7	2020	Li et al. [92]	5
3	K_2CO_3	1	2009	Mohammadi et al. [96]	10
4	KBr	1	2009	Mohammadi et al. [96]	12
5	KCl	1	2008	Mohammadi et al. [95]	13
		2	2009	Haghighi et al. [84]	3
		3	2016	Cha et al. [102]	15
		4	2017	Hu et al. [91]	27
		5	2019	Aregbe et al. [98]	23
		6	2020	Li et al. [92]	7
6	KHCOO	1	2017	Kamari et al. [88]	15
7	$MgBr_2$	1	2021	Bhawangirkar and Sangwai [101]	40
8	$MgCl_2$	1	1998	Kang et al. [89]	22
		2	2006	Atik et al. [100]	21
		3	2009	Haghighi et al. [84]	4
		4	2009	Mohammadi et al. [96]	10
		5	2017	Chong et al. [103]	6
		6	2019	Du et al. [83]	5
		7	2020	Li et al. [92]	5
9	Na_2SO_4	1	2006	Maekawa and Imai [93]	13
10	NaBr	1	2006	Maekawa and Imai [93]	27
		2	2009	Mohammadi et al. [96]	13
11	NaCl	1	1983	de Roo et al. [81]	23
		2	1991	Dholabhai et al. [82]	6
		3	1995	Maekawa et al. [94]	50
		4	2001	Jager and Sloan [58]	42
		5	2006	Maekawa and Imai [93]	30
		6	2008	Mohammadi et al. [95]	5
		7	2009	Haghighi et al. [84]	6
		8	2016	Cha et al. [102]	16
		9	2017	Hu et al. [97]	28

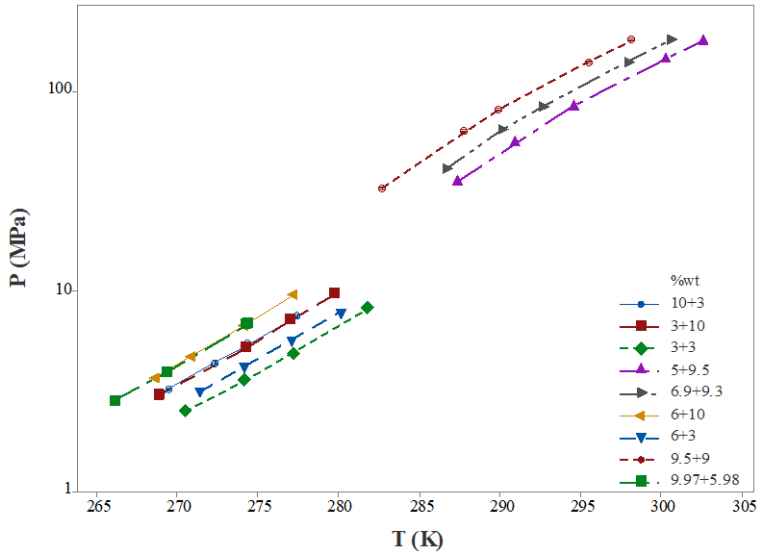
Index	Brine	Index	Year	Reference	N
		10	2019	Aregbe et al. [98]	36
		11	2019	Du et al. [83]	13
		12	2020	Li et al. [92]	7
12	NaHCOO	1	2017	Kamari et al. [88]	15
13	NaI	1	2006	Maekawa and Imai [93]	9
14	NH_4Cl	1	2016	Cha et al. [102]	21
15	$ZnBr_2$	1	2021	Bhawangirkar and Sangwai [101]	38
		Multi Salt			
1	$CaCl_2$+$MgCl_2$	1	2017	Hu et al. [85]	3
2	NaCl+$CaCl_2$	1	1991	Dholabhai et al. [82]	24
		2	2017	Hu et al. [85]	15
3	NaCl+$CaCl_2$+$MgCl_2$	1	2017	Hu et al. [85]	3
4	NaCl+KCl	1	1991	Dholabhai et al. [82]	31
		2	2017	Hu et al. [85]	6
5	NaCl+KCl+$CaCl_2$+$MgCl_2$	1	2017	Hu et al. [85]	6
6	NaCl+$MgCl_2$	1	2017	Hu et al. [85]	3

Table 2-2: The ranges for brine concentrations based on both salt and ion types.

Brine	*C*, %wt	Ion	*x*, mole fraction
$CaBr_2$	0.1 – 32	Na^+	0.0066 – 0.1167
$CaCl_2$	1.5 – 35.5	K^+	0.0064 – 0.1388
K_2CO_3	5 – 10	NH_4^+	0.0174 – 0.0361
KBr	5 – 10	Zn^+	0.0001 – 0.0139
KCl	2.6 – 40	Ca^{++}	0.0001 – 0.0820
KHCOO	5 – 20	Mg^{++}	0.0001 – 0.0323
$MgBr_2$	0.1 – 15	Cl^-	0.0038 – 0.1640
$MgCl_2$	1 – 15	Br^-	0.0002 – 0.0814
Na_2SO_4	10 – 10	I^-	0.0132 – 0.0132
NaBr	5 – 20	$HCOO^-$	0.0111 – 0.0621
NaCl	2.1 – 30	CO_3^{--}	0.0068 – 0.0143
NaHCOO	5 – 20	SO_4^{--}	0.0139 – 0.0139
NaI	10 – 10		
NH_4Cl	5 – 10		
$ZnBr_2$	0.1 – 15		



(a)



(b)

Figure 2-1: The P-T diagram for (a) NaBr and (b) NaCl+$CaCl_2$ systems.

All the single-salt systems have the same trend regarding the pressure (P) and salt concentration (C) effects on the equilibrium temperature (T). The P-T diagrams for the NaBr (5%, 10% and 20%) system is depicted in Figure 2-1 (a) as an example. This figure shows that in a constant P, an increase in the C results in a decrease in the equilibrium T. On the other hand, for a fixed C, there is a monotonic increase in T by increasing P. The same goes true for the multi-salts systems (see Figure 2-1 (b) for NaCl+$CaCl_2$). This figure suggests that the concentration of different ions has a significant influence on the phase equilibria.

2.3.2. Data quality check

It is always probable that incorrect data creep into the reported database for several reasons such as errors in the devices or human error in recording. Without a correct and clean database, the developed predictive models would not be accurate and general enough. Hence, all the data were checked for any inconsistency. In data reported by Dholabhai et al. [82], there is an inconsistent T reported as 72.22 K. Considering the T values reported for other data samples and the authors' statement for the tested T range of 264-284 K, it seems that there has been an error in recording, and the correct value might have been 272.22 K.

Additionally, the database was checked for duplicate samples. The presence of duplicates has a deleterious effect on the machine learning model's training as it cannot learn the true pattern. Table 2-3 lists the present duplicates in the gathered database. As can be seen, 5 and 5 cases were detected for pure water and NaCl systems, respectively. For the pure water, 4 cases of duplicates were reported in different studies. On the other hand, all the duplicates for NaCl brine were reported in the same study. As the duplicate values are very close to each other, the discrepancy might be due to the different methods of measurement or repeating the measurements. In this study, one of the duplicates was excluded from the database.

Table 2-3: The present duplicates in the gathered database.

Year	Author	Source	Brine	%wt of salt	P (MPa)	T (K)
1983	de Roo et al. [81]	Table 1	Pure Water	0	3.34	275.41
1983	de Roo et al. [81]	Table 1	Pure Water	0	3.34	275.95
2006	Maekawa and Imai [93]	Figure 4	Pure Water	0	4.19	277.88
2021	Bhawangirkar and Sangwai [101]	Table S1	Pure Water	0	4.19	277.8
2006	Maekawa and Imai [93]	Figure 4	Pure Water	0	5.72	280.9
2021	Bhawangirkar and Sangwai [101]	Table S1	Pure Water	0	5.72	281
2017	Kamari et al. [88]	Table 11	Pure Water	0	7.25	283.3
2021	Bhawangirkar and Sangwai [101]	Table S1	Pure Water	0	7.25	283.1
1991	Adisasmito et al. [99]	Table II	Pure Water	0	8.55	284.7
2017	Kamari et al. [88]	Table 11	Pure Water	0	8.55	284.8
2017	Hu et al. [97]	Table S1	NaCl	12	37.8	290.41
2017	Hu et al. [97]	Table S1	NaCl	12	37.8	290.44
2017	Hu et al. [97]	Table S1	NaCl	12	62	293.55
2017	Hu et al. [97]	Table S1	NaCl	12	62	293.75
1995	Maekawa et al. [94]	Figure 2	NaCl	20	6.85	270.25
1995	Maekawa et al. [94]	Figure 2	NaCl	20	6.85	270.60
2017	Hu et al. [97]	Table S1	NaCl	23	32	278.26

Year	Author	Source	Brine	%wt of salt	P (MPa)	T (K)
2017	Hu et al. [97]	Table S1	NaCl	23	32	278.35
2017	Hu et al. [97]	Table S1	NaCl	26	62	278.75
2017	Hu et al. [97]	Table S1	NaCl	26	62	278.85

2.3.3. Feature selection for modeling

Almost all the experimental works in the literature have reported C in %wt. However, it is opted to use ion's mole fraction (x) for the modeling purpose for several reasons. First of all, using C, it is not possible to correctly address the multi-salts systems. As was illustrated in Figure 2-1 (b), the effects of the presence of different salts cannot be adequately shown by C. Secondly, using ions concentration, there would be 12 characteristic features for brines, which is lower than the number of different salts, which is 15. Using a fewer number of features not only decreases the processing load but also lowers the overfitting probability. More importantly, using the ions concentration, it is possible to address the salts not present in the reported experimental data samples. Among the listed ions in Table 2-2, there are 6 cations and 6 anions. Thus, 36 different single-salt brines can be theoretically developed. As is listed in the same table, experimental data are available for only 15 single-salt systems. We will explore the capability of the developed predictive models for the single-salt systems the data for which are not provided directly. In addition to brine characteristics, P will also be used as the operational parameter to predict the equilibrium T for the CH_4-hydrate.

2.3.4. Data splitting

The gathered database includes 977 data samples after removing the duplicates. Data splitting in the ratio of 80/20 for training/testing was considered. Before splitting, 40 data samples belonging to $MgBr_2$ were set aside for additional testing. As is shown in Table 2-1, the experimental data for this brine were reported recently by Bhawangirkar and Sangwai [101]. Although there is no other reference reporting experimental data for this salt, data samples are available containing both Mg^{++} and Br^{-} ions for other salts. As additional testing, if the developed model provides an acceptable prediction for $MgBr_2$, it would be able to be used for any hypothetical salt the ions of which are available in the database. The remaining 937 data samples were split into training and testing subsets in the ratio of 80/20. A 5-fold Cross-Validation (CV) was used to validate the developed models.

2.4. Result and discussion

2.4.1. Model development

In this section, the steps of model development including hyperparameter optimization and model fitting are introduced. The Scikit-Learn library [79] developed for Python is used in this regard. The road map is as follows. Three different predictive models of MLP, DT, and ET are considered to develop. In the first step, the hyperparameters for each model are optimized. It is worthwhile mentioning that only the main ones are subjected to optimization. Then, the models are fitted several times due to the stochastic nature of the models. The best one for each model type is saved for the accuracy analysis. The accuracy will be checked using both graphical and statistical methods for the training, testing, and additional testing datasets.

2.4.1.1. Hyperparameter optimization

Hyperparameters denote the external shape of a model and should be determined before the training step. On the other hand, the model's parameters are determined during the training step. The best set of hyperparameters guarantees the best performance of the model for the given dataset. It goes without saying that just like all other actions, hyperparameter optimization is done just using the training dataset. Table 2-4 shows the determined hyperparameters as well as the searching space for each model. As the search space for the MLP was discrete, an exhaustive search was performed for networks with one hidden layer with neurons in the range of (1,)-(10,) and networks with two hidden layers with neurons in the range of (1,1)-(10,10). Among the 110 assessed cases, the MLP with two hidden layers with (10,10) neurons were determined to be the best structure. On the other hand, a randomized search with 20 iterations was used for DT and ET methods. The best 5-fold CV Root Mean Squared Error (RMSE) values are listed in the table. It is worthwhile mentioning the reported error values are for the standardized T.

Table 2-4: The determined hyperparameters as well as the searching space for each model.

	Hidden_layer-size	max_depth	max_features	n_estimators	Best CV RMSE
Search Interval	(1,)—(10, 10)	5—50	0.1—1	5—50	—
MLP	(10, 10)	—	—	—	0.1993
DT	—	43	0.7112	—	0.2735
ET	—	21	0.7580	43	0.1842

2.4.1.2. Model fitting

The scaled values were used for MLP, DT, and ET models. Figure 2-2 depicts the box-plot for the 5-fold CV RMSE acquired from 20 trials for MLP, DT, and ET models. As is obvious, ET yielded the best results with a mean value equal to 0.19. Regarding the accuracy, after ET are MLP and DT models with respective mean values of 0.23 and 0.29. All the RMSE values for ET are very close to each other. The closer error values to each other for different random states, the more stable the model. As each trial uses a different random state, the yielded results are not the same. This shows the stability of this model.

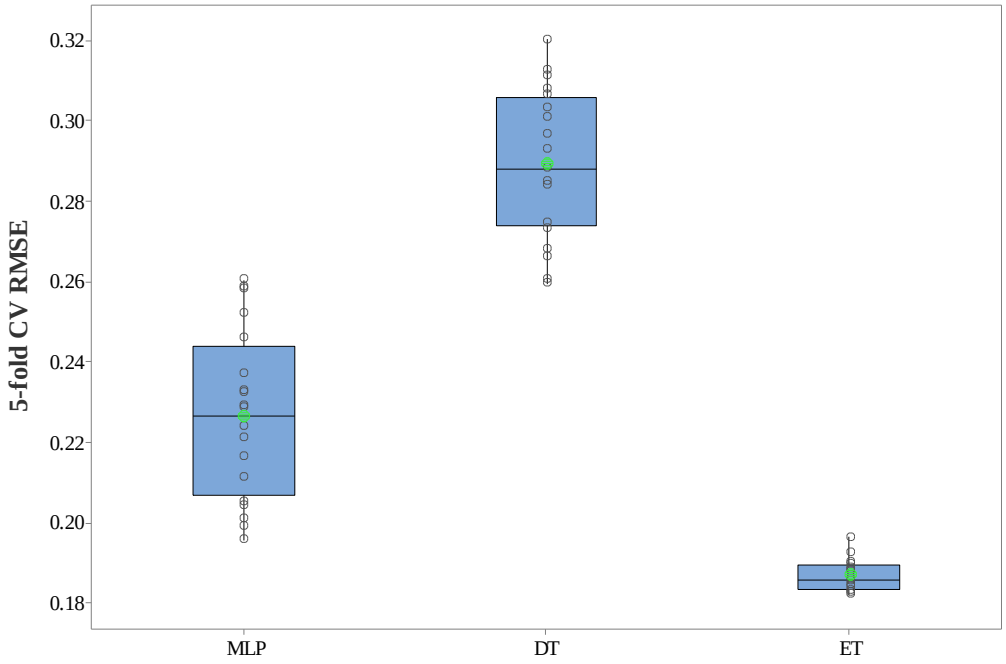


Figure 2-2: The box-plot for the 5-fold CV RMSE acquired from 20 trials for MLP, DT, and ET models.

2.4.2. Accuracy assessment

Different statistical parameters can be used to show the error of predictions. In this study, Residual (Res_i), Coefficient of determination (R^2), RMSE, Average Relative Deviation (ARD), and Average Absolute Relative Deviation (AARD). The formulations of the aforementioned parameters are as follows:

$$Res_i = t_i - y_i \tag{2-1}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Res_i)^2}{\sum_{i=1}^{n}(\bar{t} - t_i)^2} \tag{2-2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Res_i)^2} \tag{2-3}$$

$$ARD = \frac{100}{n} \times \sum_{i=1}^{n}\frac{Res_i}{t_i} \tag{2-4}$$

$$AARD = \frac{100}{n} \times \sum_{i=1}^{n}\frac{|Res_i|}{t_i} \tag{2-5}$$

where y and t are the model's output and the target value, respectively.

Table 2-5 lists the statistical parameters denoting the accuracy of the developed models in this study. The statistical parameters are calculated for the original T, not the scaled ones. The models must be compared using the testing dataset. In this table, the best values are bolded. As is shown, ET has yielded the best results by an RMSE equal to 0.6248 for the testing dataset. Except for ARD, the ET has exhibited the best results for all other parameters. The other important point is the difference between the training and testing errors. For MLP, the errors for the training and testing datasets are close to each other. On the other hand, DT shows a noticeable difference, which is indicative of overtraining. In other words, the developed DT model has failed in showing good performance as it did for the training data. This problem is addressed in the ET model, which is comprised of several individual DTs. By taking the average of individual estimators, not only has the ET model provided excellent results for the training dataset but also has shown very good performance for the testing dataset.

Table 2-5: The statistical parameters denoting the accuracy of the developed models in this study.

Model	Dataset	R^2	RMSE	ARD (%)	AARD (%)	N
MLP	Training	0.9825	0.9978	0.0031	0.1943	749
	Testing	0.9837	0.9815	-0.0058	0.2263	188
DT	Training	1.0000	0.0000	0.0000	0.0000	749
	Testing	0.9621	1.4968	0.0431	0.3792	188
ET	Training	1.0000	0.0380	0.0000	0.0068	749
	Testing	**0.9934**	**0.6248**	**0.0026**	**0.1436**	188

Figure 2-3 illustrates the comparison between predicted and experimental T for the testing dataset. Subplots (a) and (b) shows the cross and residual plots, respectively. In both subplots, the x-axis depicts the experimental T. For a perfect prediction, all the data samples lie on the 45° line and zero line for the cross and residual plots, respectively. As one can see, the ET has provided the best results. Both statistical and graphical error analyses corroborate that the proposed ET model is well capable of predicting unseen data samples during the training phase.

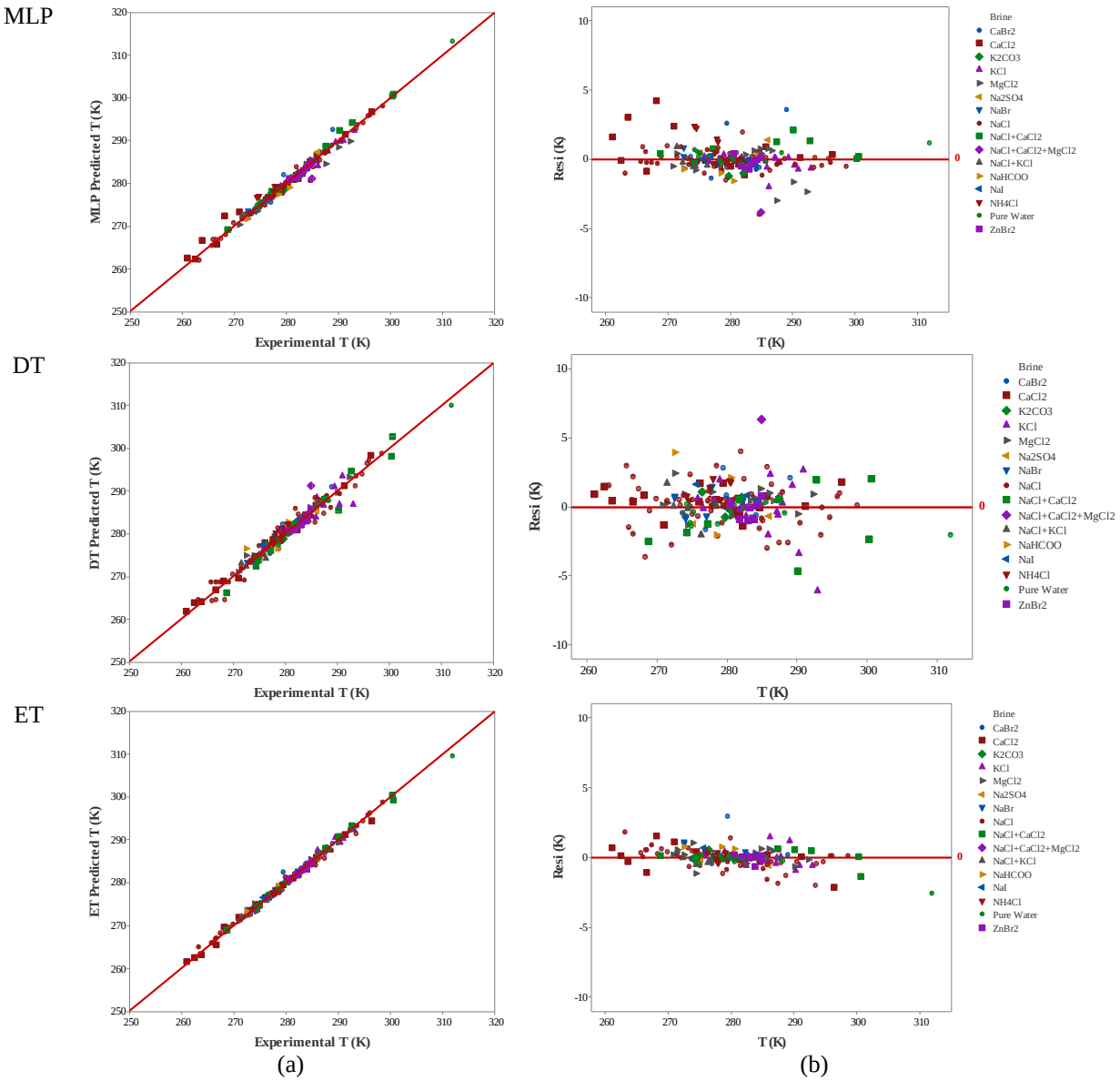


Figure 2-3: comparison of predicted and experimental T for the testing dataset. (a) cross-plot and (b) Residual plot.

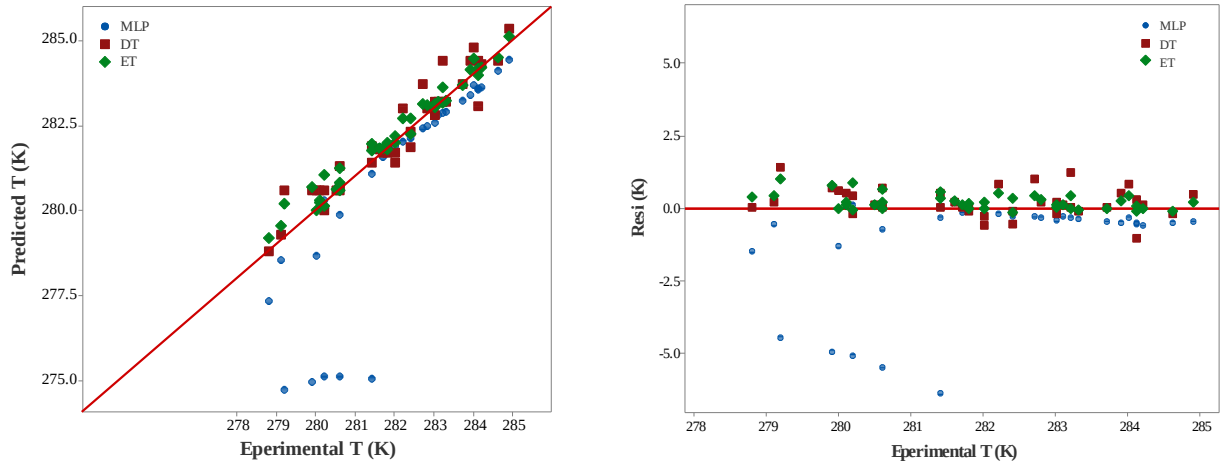
2.4.3. Additional testing

As was mentioned in section 2.3.4, the data samples belonging to the $MgBr_2$ system were set aside for additional testing before splitting data for training and testing. The reason for selecting this specific system was the fact that there were no instances of this brine in multi-salts systems and also there were several occurrences of its comprising ions in other salts. Thus, it was expected that a well-trained model should be able to provide an adequate result for this system. Table 2-6 lists

the statistical parameters. The best values are bolded. This table suggests that only MLP and ET have provided good results for $MgBr_2$. Figure 2-4 shows both cross and residual plots. The residual plot shows that although having overall acceptable predictions, MLP is deficient for exhibiting a trend in residuals. For the first and second halves of the T, MLP tends to under and over-predict, respectively. All in all, the ET provided the best results for the additional testing as it did for training and testing datasets.

Table 2-6: Statistical error analysis of different proposed models for the $MgBr_2$ system.

Model	R^2	RMSE	MPE (%)	MAPE (%)	N
MLP	-0.4026	1.9308	**-0.3321**	0.3488	40
DT	0.8997	0.5163	0.0726	0.1324	40
ET	**0.9534**	**0.3520**	0.0788	**0.0906**	40



(a) (b)

Figure 2-4: Comparison among different models for predicting equilibrium T for $MgBr_2$ system (a) cross-plot and (b) residual plot.

2.4.4. Comparison with literature

The results acquired in this study are compared with the models published recently in 2021 by Xu et al. [29]. Table 2-7 shows the comparison between the models developed in this study and literature models regarding the RMSE for the testing dataset. The best values are bolded. As one can see the mean RMSE for the developed ET and GBT proposed by Xu et al. [29] are close to each other with a slight superiority of the latter. However, the models developed in this study cover a much wider range of different salts. To be more specific, 15 different salts are used in this study

compared with 4 distinct salts used by Xu et al. [29] Regarding the different ions, this study addresses 6 cations and 6 anions. On the other hand, the study of Xu et al. [29] includes only 4 cations and 1 anion. The comparison reveals that the developed ET model has provided a comparable accuracy with the GBR model proposed by Xu et al. [29] but for a wider range of brines. Most of the time, the prediction models based on machine learning methods do not provide an explicit equation for predicting equilibrium. However, decision tree–based models allow direct inspection of prediction rules, improving interpretability. Additionally, implementing a web app or web-based user interface allows us to illustrate the application of these models.

Table 2-7: Comparison between the models developed in this study and literature models regarding the RMSE for the testing dataset.

	Xu et al. [29]					**This study**		
Brine	**MLR**	**k-NN**	**SVM**	**RF**	**GBR**	**MLP**	**DT**	**ET**
$CaBr_2$	4.54	4.73	2.6	1.08	1.05	1.33	1.14	**0.83**
$CaCl_2$	—	—	—	—	—	1.47	0.99	**0.77**
$CaCl_2$+$MgCl_2$	17.94	12.53	1.02	4.3	**0.42**	—	—	—
K_2CO_3	—	—	—	—	—	0.94	0.84	**0.32**
KCl	4.22	1.67	1.5	1.47	0.66	0.63	2.08	**0.60**
KCl+$CaCl_2$	4.81	1.9	1.46	0.26	**0.06**	—	—	—
$MgBr_2$	—	—	—	—	—	1.93	0.52	**0.35**
$MgCl_2$	4.54	3.71	3.32	1.18	0.71	1.08	0.90	**0.49**
NH_4Cl	—	—	—	—	—	1.35	1.28	**0.31**
Na_2SO_4	—	—	—	—	—	0.84	0.89	**0.37**
NaBr	—	—	—	—	—	**0.45**	0.80	0.55
NaCl	4.24	1.66	2.44	0.81	0.7	0.78	1.53	**0.66**
NaCl+$CaCl_2$	3.98	2.19	1.88	0.92	**0.57**	0.99	2.30	**0.56**
NaCl+$CaCl_2$+$MgCl_2$	6.62	15.57	**0.02**	2.87	0.96	3.82	6.30	0.27
NaCl+KCl	4.85	2.29	2.94	0.58	0.91	0.57	1.33	**0.41**
NaCl+KCl+$CaCl_2$+$MgCl_2$	4.98	2.23	0.84	0.57	**0.1**	—	—	—
NaHCOO	—	—	—	—	—	1.16	2.83	**0.69**
NaI	—	—	—	—	—	**0.03**	1.23	0.50
Pure Water	4.73	**0.23**	2.73	0.38	0.59	0.49	0.92	0.78
$ZnBr_2$	—	—	—	—	—	0.41	0.66	**0.26**

2.5. Conclusions

The purpose of this study was to develop machine learning-based models to accurately predict the methane-hydrate formation temperature in presence of different brines. In this regard, firstly, a comprehensive database was gathered from the open literature. It covered a wide range of pressure and temperature and 15 different brines. Data were checked for unreliable data and duplicates. It was opted to use ions' mole fractions to represent the brines so that the developed models would be usable for any possible combination of ions present in the database. Before data splitting into

training and testing, $MgBr_2$ samples were set aside for additional testing. The remaining data were split in the ratio of 80/20 for respective training and testing. Three different models of MLP, DT, and ET were used. Firstly, the hyperparameters were optimized using an exhaustive search for the MLP model and a randomized search for DT and ET. Then the models were fitted using the training dataset. The ET model showed the best performance with an RMSE equal to 0.6248 K for the testing dataset. The order of model based on their accuracy was as ET > MLP > DT for the testing data set. Additionally, ET showed the best stability during fitting for 20 different random states. For the additional testing, ET resulted in an RMSE equal to 0.3520, which shows the capability of this model to be used for any brine that can be developed with ions available in the database. All the models also showed acceptable performance for the additional testing. The models were compared with the most recent ones proposed in the literature. The comparison showed a comparable level of accuracy with the developed models covering a much wider range of brines and ions. The findings of the present study can be used as a powerful tool for predicting the methane-hydrate formation conditions in the pure water, single-salt brines, and multi-salt brines.

2.6. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rser.2022.113103.

Chapter 3: Insights into predicting equilibrium conditions of clathrate hydrates of methane + water-soluble hydrate former

3.1. Introduction

Gas hydrates are non-stoichiometric crystalline structures wherein small gas molecules trapped within water cages, held together by hydrogen bonds. These compounds usually form under high pressure and low temperature conditions, with water acting as the host and the gas molecules as guests [2]. Structure I (sI), Structure II (sII), and Structure H (sH) represent three prevalent structures of gas hydrates. The size of the gas molecules trapped within the water cages, known as van der Waals radii, determines the formation of each structure [104]. Typical gases that create sI hydrates are methane (CH_4), ethane, carbon dioxide (CO_2), and xenon. sII hydrates can be formed by natural gas mixtures like propane and cyclopentane. For sH hydrates to stabilize, they require a combination of two gas molecules, such as CH_4 and methylcyclohexane [105].

In addition to clathrate hydrates, there are semiclathrates, which are crystalline structures involving water and hydrophobic molecules. In semiclathrates, ions from the hydrophobic molecules replace some of the water molecules and become incorporated into the clathrate cages. This means that the molecules forming semiclathrates act as both hosts and guests. Compounds like tetra-n-butylammonium bromide and tetra-n-butylammonium fluoride can produce semiclathrate hydrates with guest gases such as CH_4 and CO_2 [106,107].

While hydrates are often viewed as problematic within the oil and gas sector [108,109], their characteristics make them suitable for various innovative applications. This includes their use in gas separation processes, where the target gas selectively partitions between the hydrate crystal phase and the gas phase [6,7]. Furthermore, gas hydrates find application in other processes including seawater desalination [12,13], the capture of greenhouse gases [8,9], and secondary refrigeration [10,11]. Moreover, Gas hydrates have a special ability: they can store gas at a really high ratio, about 170 v/v (volume of gas to volume of hydrate) [2]. This makes them great for storing natural gas, known as called Solid Natural Gas (SNG) technology, and for transporting it around the world efficiently [50]. SNG technology effectively integrates safety, environmental sustainability, and economic feasibility, making it a great choice for large-scale gas storage [51].

Scientists are currently focusing on using SNG technology to store CH_4, which is one of the most common natural gases [110].

Hydrate formation typically needs high pressure [2]. When certain organic compounds, referred to as water-soluble or water-insoluble hydrate formers, are used, they either reduce the pressure needed for hydrate equilibrium at a particular temperature or increase the temperature required at a specific pressure [111–113]. The selection of these water-soluble or water-insoluble compounds takes into account factors such as economics, safety, and environmental impact [2,112]. Water-soluble organic hydrate formers may include cyclic ethers and ketones, those that do not integrate into the hydrate cavities, such as tetrahydrofuran (THF) [114], 1,3-dioxalane [115], 1,4-dioxane, and acetone [105], and similar substances, typically contribute to the formation of sII clathrate hydrates, whether gas is present or not [112,116]. When organic compounds are added at concentrations up to 6 mol%, approximately corresponding to the stoichiometric ratio of large sII cages to water (1/17), they decrease the equilibrium pressure of hydrate formation. However, adding concentrations beyond 6 mol% results in a gradual increase in equilibrium pressures [116,117].

To date, numerous experimental investigations have been carried out on hydrates involving both water-soluble and water-insoluble organic compounds. However, there have been relatively few theoretical studies aimed at modeling the equilibrium conditions of these compounds, especially those that are water-soluble hydrate formers [117].

The available thermodynamic models in the literature for predicting gas hydrate phase equilibria in the presence of formers typically rely on the equality of fugacities/chemical potentials of components in the present phases [118,119]. The solid solution theory of van der Waals and Platteeuw, known as the vdW-P model [120], is employed to assess the fugacity or chemical potential of water within the hydrate phase. When formers are dissolved into water, it results in a significant increase in the polarity of the aqueous phase. It is widely recognized that only a few models can reliably predict the water activity coefficient [121]. Improving thermodynamic models to predict gas hydrate dissociation conditions with different formers necessitates access to equilibrium solubility data for these formers in water. Unfortunately, such data is notably lacking in existing literature, thus limiting the effectiveness of thermodynamic models [26]. While the UNIFAC group contribution model can deal with this challenge, employing such methods can be

complex, particularly when dealing with molecules lacking readily available parameters [122]. Furthermore, over time, an increasing array of formers becomes accessible. Nevertheless, existing thermodynamic models relying on the vdW-P model [120] are limited to represent gas hydrate dissociation conditions in the presence of specific formers, their applicability to other hydrate formers has not yet been evaluated [121]. For instance, Illbeigi et al. [112] developed a model to predict the equilibrium conditions of clathrate hydrates formed by cyclopentane, cyclohexane, acetone, or 1,4-dioxane + CH_4 (sII), and 1,1-dimethylcyclohexane, 2,2-dimethylpentane, methylcyclohexane, or cis-1,2-dimethylcyclohexane + CH_4 (sH). Their proposed thermodynamic model employs the van der vdW-P model for modeling the hydrate phase. The gas phase is modeled using the Peng–Robinson (PR) and Soave–Redlich–Kwong (SRK) equations of state with random mixing rules, while the hydrocarbon and aqueous phases are modeled using the UNIFAC method.

Furthermore, the development of a model for predicting gas hydrate dissociation conditions in the presence of formers using thermodynamic models necessitates the utilization of many adjusted parameters. Typically, significant efforts are required to establish an appropriate relationship for fitting experimental data within these models [123]. Additionally, thermodynamic-based models usually perform well within the ranges on which they were developed, and they may not provide high accuracy beyond those ranges. Moreover, if the thermodynamic former is changed, recalculating the adjusted parameters is necessary as the developed models may not predict with good accuracy. However, models based on machine learning (ML) algorithms offer valuable features such as the identification of patterns in the data, the ability to model nonlinear complex systems, and high accuracy for the utilized experimental data [30].

Many studies emphasize the role of ML in gas hydrate research. Lin et al. [124] used ML and molecular dynamics (MD) simulations to study THF hydrates, demonstrating that strength varies with ice content, strain rate, and temperature, effectively predicted by extreme gradient boosting models. Zhang et al. [125] employed ML and MD to investigate how occupancy rates and guest molecule distributions impact CO_2 hydrate's mechanical properties, highlighting the crucial insights for CO_2 geostorage. Lin et al. [126] applied ML and MD to show that increasing oxidation levels on graphene oxide surfaces significantly influences methane hydrate formation, suggesting optimal levels for enhanced gas storage and transport. Shi et al. [127] utilized ML and MD to

analyze the effects of varying strain levels on methane hydrate growth, identifying structural changes and new hydrate forms, with a long short-term memory network highlighting the critical role of strain.

With a focus on the gas hydrate equilibrium conditions, Mohammadi and Richon [123,128] made the initial efforts in this area. In 2008, Mohammadi and Richon [123] developed a model using the feed-forward ANN method, employing a modified Levenberg−Marquardt optimization algorithm. This model aimed to predict the solubility of a pure hydrocarbon hydrate former in pure water under gas hydrate equilibrium conditions. Their study included data for three hydrocarbon hydrate formers, namely CH_4, ethane, and propane. Their findings demonstrated an acceptable agreement between the predicted and experimental values. Subsequent to this research, in 2010, Mohammadi and Richon [128] a proposed a model based on the feed-forward ANN algorithm to predict hydrate dissociation conditions for the hydrogen+water and hydrogen+ THF +water systems. The model provided acceptable predictions for hydrate dissociation conditions when THF concentrations in the aqueous phase were below its stoichiometric concentration (approximately 0.059). These two innovative research have inspired researchers to pursue further studies in this field.

Eslamimanesh et al. [26] presented two molecular models to predict CH_4 hydrate dissociation conditions in the presence of water-insoluble promoters using the quantitative structure–property relationship (QSPR) strategy. The least square support vector machines (LSSVM) and the genetic-algorithm-based multivariate linear regression (GA-MLR) models were employed to determine the model parameters. Their results exhibited acceptable agreement between the predicted and experimental values, with the QSPR-LSSVM nonlinear model showing higher accuracy compared to the QSPR-MLR linear model. In 2013, Ilani-Kashkouli et al [121] employed the LSSVM method to assess the reliability of experimental data reported for the dissociation conditions of sII and sH hydrates of four different gas species, namely, CH_4, CO_2, H_2, and N_2, in the presence of water-soluble and insoluble organic hydrate former components. Babakhani et al. [129] contributed to this field by developing a feed-forward multi-layer ANN to predict the hydrate formation pressure of hydrate binary mixture systems, including at least one hydrocarbon. Their model comprised one input layer, one hidden layer with seven neurons, and one output layer. Various physical properties such as critical temperature, critical pressure, molecular weight, mole fraction, and hydrate dissociation temperature served as inputs. When comparing their model to

the Colorado School of Mines Hydrate (CSMHYD) software, Babakhani et al. [129] found that their model yielded more accurate results. Mesbah et al. [130] continued this trend by developing a multi-layer perceptron (MLP) ANN model to predict the dissociation pressure of semi-clathrate hydrates within CO_2+N_2, CO_2+H_2, CO_2+CH_4, and pure CO_2 systems, considering various concentrations of tera-n-butyl ammonium bromide (TBAB) in aqueous solutions. Input parameters included the weight percent of TBAB, mole fraction of gas species, and temperature. Their findings demonstrated the effectiveness of the ANN model in accurately predicting semi-clathrate hydrate dissociation pressure. Additionally, Sinehbaghizadeh et al. [131] developed a multilayer perceptron feed-forward ANN model and an LSSVM model to predict the dissociation temperature of various four-phase sII double hydrate systems. The performance of the ANN model was observed to surpass that of the LSSVM model in determining the hydrate dissociation temperature. In a recent investigation, Pornaroontham et al. [132] developed a machine learning (ML)-based model to analyze the equilibrium conditions of CH_4 hydrate-forming systems with non-ionizing water-soluble organic compounds. They utilized five algorithms: multiple linear regression, polynomial regression, decision tree regressor, random forest, and categorical boosting machine regressor. Their research provided promising insights into selecting suitable hydrate formers.

This study aims to develop a comprehensive ML model for predicting equilibrium conditions in CH_4 hydrate systems involving a broad range of water-soluble hydrate formers. We will begin by organizing and refining an extensive dataset from existing literature, thereby updating and expanding previous research in this domain. The database will undergo meticulous scrutiny using standard statistical methods before being used for modeling.

This research sets itself apart from similar endeavors in several key ways. Firstly, it leverages a more extensive and diverse dataset, enhancing the models' generalizability and providing a richer learning environment for the ML algorithms. Additionally, inspired by Pornaroontham et al. [132], this study will explore a variety of non-linear transformations beyond the square transformation previously employed. This approach aims not only to develop a more accurate model but also to address issues related to extreme data samples innovatively.

A significant innovation in our approach is the implementation of a rigorous model testing procedure. While Pornaroontham et al. [132] employed stratified data splitting to maintain a

consistent ratio of samples across training and testing datasets for each former, our study proposes a novel splitting strategy. By focusing on the diversity of formers rather than the sheer volume of data samples, we aim to ensure that the predictive model remains versatile and applicable to formers beyond those explicitly included during the model training phase. Furthermore, based on the ML modeling results, we will employ Shapley Additive Explanations (SHAP) values to analyze the impact of important features of formers and operational parameters on equilibrium temperature. This analysis is crucial for identifying molecular features and operational parameters that contribute to optimal performance in the use of thermodynamic formers, where the formers can be efficient and capable of shifting the equilibrium temperature to higher levels.

3.2. Database

The database, originally containing 847 hydrate phase equilibrium data points involving 34 various thermodynamic formers as provided by Pornaroontham et al. [132], served as the foundation. Indeed, experimental data on the three-phase (water–hydrate–vapor (L_W–H–V)) equilibrium temperatures and pressures for CH_4 hydrate in pure water with varying concentrations of water-soluble formers [116,117,133–153] have been employed.

Table 3-1 presents the details of the gathered experimental data. This table represents a comprehensive database compiled from the reviewed literature, outlining the experimental conditions of a diverse set of chemical compounds referred to as "formers." Each entry in the table includes the common name of the compound ("Former"), its structural representation in Simplified Molecular-Input Line-Entry System (SMILES) [154] notation ("SMILES"), and its molar fraction range ("mol_fr") within the studied system. The table also specifies the range of pressures ("P_(MPa)") and temperatures ("T_(K)") under which the compounds were analyzed. These ranges indicate the conditions at which the compounds exhibit certain thermodynamic or physical behaviors, as determined through experimental measurements. Additionally, the table provides references ("Reference") listing the source of data samples.

Table 3-1: Database of chemical compounds: physicochemical properties and experimental conditions.

Former	SMILES	mol_fr	P_(MPa)	T_(K)	Reference(s)
1,2-Ethanediol	C(CO)O	0.020 - 0.350	2.79 - 18.30	247.40 - 284.70	[144,145]
1,3-Dioxolane	C1COCO1	0.010 - 0.200	0.64 - 14.05	282.60 - 299.40	[117,141]
2,5-Dihydrofuran	C1C=CCO1	0.007 - 0.043	6.25 - 12.17	291.60 - 300.30	[133]
2-Methylpiperidine	CC1CCCCN1	0.029 - 0.029	4.82 - 7.74	288.40 - 294.10	[150,152]
3-Methylpiperidine	CC1CCCNC1	0.029 - 0.029	4.18 - 7.16	286.00 - 293.50	[150,152]
4-Methylpiperidine	CC1CCNCC1	0.029 - 0.029	5.50 - 7.92	284.00 - 289.40	[150,152]
Acetone	CC(=O)C	0.010 - 0.477	1.03 - 11.30	273.60 - 292.30	[134,143,149]
Arginine	C(C[C@@H](C(=O)O)N)CN=C(N)N	0.011 - 0.011	4.70 - 9.65	278.60 - 285.40	[138]
Cyclopentanone	C1CCC(=O)C1	0.011 - 0.036	7.02 - 12.99	291.20 - 298.20	[137]
DI(Hydroxyethyl)ether	C(COCCO)O	0.041 - 0.145	4.94 - 14.50	265.00 - 283.20	[144]
Diacetone alcohol (4-Hydroxy-4-methyl-2-pentanone)	CC(=O)CC(C)(C)O	0.017 - 0.062	6.99 - 13.02	278.80 - 287.00	[137]
Dioxane	C1COCCO1	0.010 - 0.300	1.06 - 14.10	275.80 - 296.50	[116,149]
DL-Alanine	CC(C(=O)O)N	0.022 - 0.022	4.61 - 9.55	277.60 - 284.30	[138]
DL-Proline	C1CC(NC1)C(=O)O	0.017 - 0.017	4.55 - 9.82	277.60 - 284.90	[138]
Ethanol	CCO	0.030 - 0.400	2.14 - 39.80	257.60 - 290.50	[148]
Glycerol	C(C(CO)O)O	0.077 - 0.164	4.51 - 64.70	266.50 - 293.50	[135]
Glycine	C(C(=O)O)N	0.012 - 0.057	4.55 - 9.98	276.50 - 285.20	[138]
Hexamethyleneimine	C1CCCNCC1	0.029 - 0.029	3.49 - 7.46	283.00 - 292.30	[150,152]
Isopropyl alcohol (2-Propanol)	CC(C)O	0.006 - 0.070	2.20 - 11.00	274.00 - 287.10	[147]
Methanol	CO	0.024 - 0.360	2.62 - 10.79	244.10 - 285.10	[145]
N-Methylpiperidine	CN1CCCCC1	0.029 - 0.029	4.00 - 6.66	287.90 - 294.30	[150,152]
Piperidine	C1CCNCC1	0.056 - 0.056	3.53 - 8.53	285.70 - 292.00	[152]
Propanol	CCCO	0.006 - 0.070	2.60 - 11.00	274.40 - 286.60	[147]
Propylene glycol	CC(CO)O	0.026 - 0.191	3.03 - 56.60	258.20 - 290.10	[135]
Propylene oxide	CC1CO1	0.030 - 0.030	1.92 - 8.76	284.00 - 294.20	[149]
Pyrazine	C1=CN=CC=N1	0.056 - 0.056	3.41 - 8.37	276.40 - 285.30	[136]
Pyridazine	C1=CC=NN=C1	0.056 - 0.056	4.39 - 7.37	273.90 - 280.70	[136]
Pyrimidine	C1=CN=CN=C1	0.056 - 0.056	2.55 - 6.77	272.20 - 282.00	[136]
Pyrrolidine	C1CCNC1	0.056 - 0.056	1.86 - 9.65	282.80 - 295.60	[142,152]
Serine	C([C@@H](C(=O)O)N)O	0.019 - 0.019	4.56 - 9.80	277.70 - 285.00	[138]
Tert-Butanol	CC(C)(C)O	0.023 - 0.196	1.77 - 10.36	276.70 - 292.10	[140,151]
Tetrahydrofuran	C1CCOC1	0.005 - 0.101	0.33 - 14.05	277.90 - 306.20	[117,139,142,146,149,153]
Tetrahydropyran	C1CCOCC1	0.010 - 0.010	2.05 - 14.05	288.30 - 302.60	[117]

Former	SMILES	mol_fr	P_(MPa)	T_(K)	Reference(s)
Triethylene glycol	C(COCCOCCO)O	0.029 - 0.107	4.64 - 12.52	266.30 - 284.60	[144]

To enhance comprehension of the chemical architecture of each examined former, Figure 3-1 presents the 2D molecular structures of each compound. For clarity and visual distinction, nitrogen and oxygen atoms are depicted in blue and red, respectively, within these representations. This figure highlights the diverse array of functional groups present in the formers, including but not limited to acids, alcohols, and ethers, highlighting the varied and complex nature of these compounds.

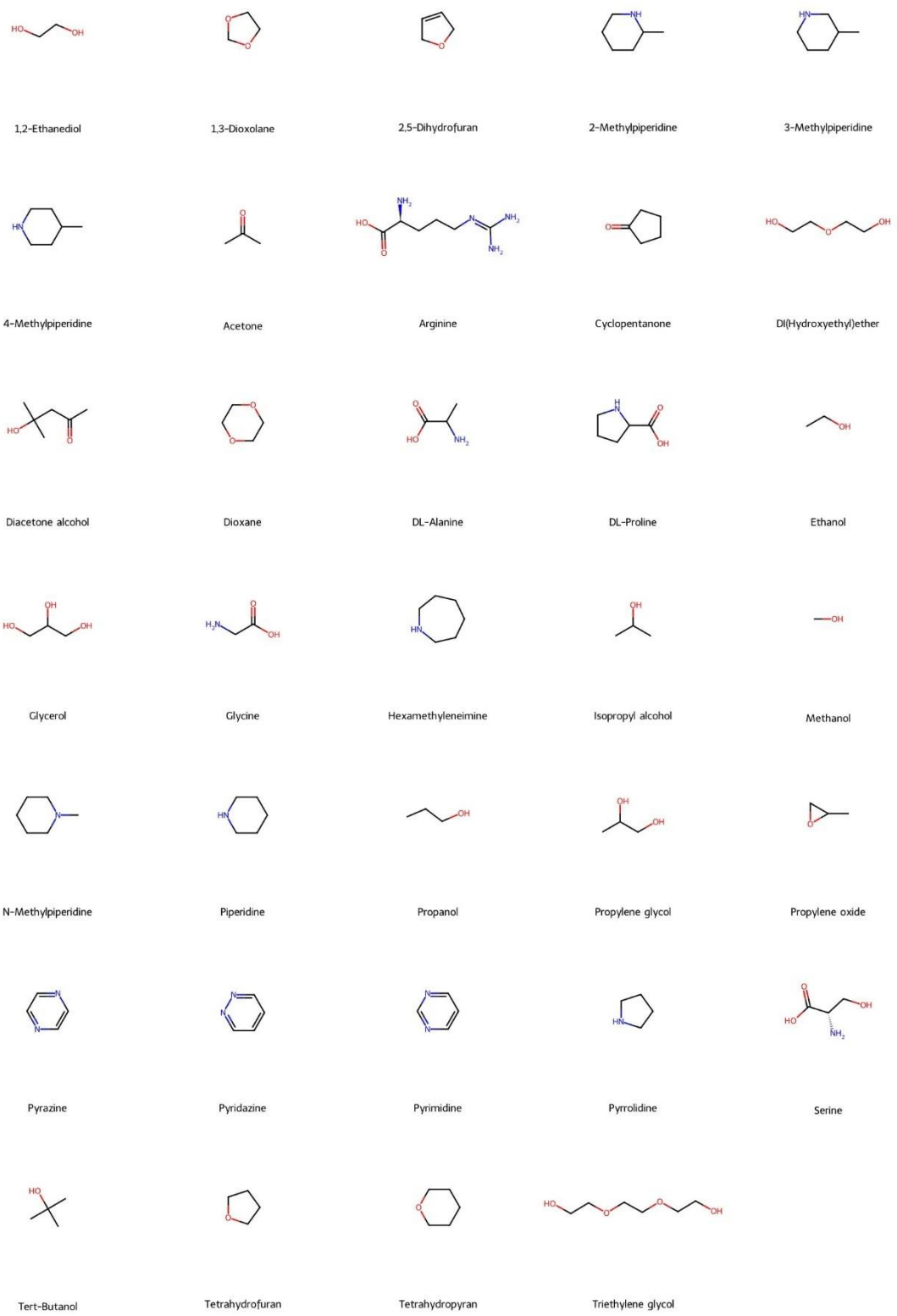


Figure 3-1: 2D molecular structures of each former included in the gathered database.

3.3. Methodology

The objective of this research is to precisely predict the dissociation temperature of CH_4 hydrate systems across a diverse range of water-soluble hydrate formers, using a minimal number of input variables. These variables comprise molecular descriptors derived from the chemical structures of various formers, along with operational parameters such as mole fraction and pressure. We employ a variety of predictive models, including Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Multi-Layer Perceptron (MLP). After selecting the final model, the SHAP approach is employed to analyze the impact of important molecular features of formers and operational parameters on equilibrium temperature, contributing to desirable performance in the use of thermodynamic formers.

Our goal is to identify a subset of the most statistically effective molecular descriptors among all available variables that can predict hydrate dissociation temperature with minimal deviation from experimental values. Furthermore, based on the modeling results, we gain new insights into the utilization of thermodynamic formers through the interpreting SHAP method, which involves examining the effect of molecular features of formers on equilibrium temperature. We detail the employed models, molecular descriptors, data splitting method, feature transformation, feature filtering, feature scaling, and SHAP method in subsequent sections.

3.3.1. Models

This study investigates different predictive models, including DT, RF, SVM, and MLP. The models are implemented using the Scikit-Learn library in Python [79].

The DT model is a ML algorithm suitable for classification and regression tasks that does not rely on any predetermined parameters. It functions by recursively splitting the dataset into subsets based on input variable values, forming a tree-like structure of decision rules. Internal nodes in this tree denote decisions based on features, while leaf nodes indicate predicted outcomes. Known for their simplicity and transparency, DTs are valuable for visual interpretation. However, excessive depth in DTs can lead to overfitting and reduced interpretability, so balancing tree depth is important for optimal performance. Thus, DTs are conceptualized as "white-box" models. In contrast, the RF model is an ensemble learning approach based on DTs. It creates multiple decision

trees during training and combines their predictions by calculating the average for regression tasks or using a majority vote for classification. This ensemble strategy addresses the overfitting tendency of individual DTs. By combining the results from multiple trees, RF enhances the model's ability to generalize and become more robust, lowering the risk of capturing noise. Known for its enhanced predictive accuracy and stability, RF is a powerful tool in different ML applications.

The SVM model is a flexible and powerful ML algorithm primarily used for both classification and regression problems. Unlike DT and RF, SVM finds the most suitable hyperplane in a high-dimensional space to effectively divide data points into different classes or predicts continuous values. SVM is especially efficient for complex datasets with non-linear decision boundaries. Unlike DT and RF, SVM functions as a black-box model, as its fitting process is not explicitly interpretable. Similarly, the MLP model, a type of artificial neural network, is employed for different ML tasks. It comprises multiple layers of interconnected neurons and is exceptional at identifying complex patterns within data. MLP is especially proficient at handling tasks with non-linear relationships, but it is also a "black-box" model due to its lack of easy interpretability. SVM and MLP are classified as distance-based models, as they depend on the relationships between data points in a multi-dimensional space. Additional insights into these predictive techniques are available in the literature [79,80].

3.3.2. Molecular descriptors

In the computational analysis of chemical structures, a series of molecular descriptors were computed to characterize the physicochemical properties and structural features of a set of formers represented by their SMILES notation. Initially, each molecular structure was converted from its SMILES format to an RDKit molecule object. Hydrogen atoms were then explicitly added to the molecular structures to accurately represent the molecular geometry, followed by the generation of 3D conformers using the ETKDG method. The 3D conformers were subsequently optimized using the Universal Force Field (UFF) to attain a low-energy, physically plausible conformation, ensuring that the computed descriptors accurately reflect the molecular properties in a stable state.

A comprehensive set of molecular descriptors was computed to capture various aspects of molecular structure and reactivity, following the recommendations of Pornaroontham et al. [132].

As listed in Table 3-2, these descriptors include molecular weight and fundamental counts of atoms, bonds, and specific atom types, providing a basic quantitative insight into the molecular composition. Additionally, topological and geometrical descriptors were calculated to describe the molecular shape, size, and the potential for intermolecular interactions. Notably, the Solvent-Accessible Surface Area (SASA) was calculated using the Shrake-Rupley algorithm [155] with a probe radius approximating that of a water molecule, providing an estimation of the molecule's exposed surface area in an aqueous environment. This parameter, alongside the calculated polar and apolar surface areas, offers a detailed understanding of the molecule's hydrophilic and hydrophobic regions, which are crucial for predicting solubility, permeability, and reactivity in systems.

Table 3-2: Molecular descriptors used in this study, employing the RDKit library [156]**.**

Index	Descriptor	RDKit Function	Description
1	mw	Descriptors.MolWt(mol)	The molecular weight of a molecule is determined by adding together the atomic weights of all the atoms it contains.
2	numAtom	mol.GetNumAtoms()	Total number of atoms in the molecule, including hydrogens if added.
3	numRT	Descriptors.NumRotatableBonds(mol)	Number of rotatable bonds: A rotatable bond is any single bond not part of a ring and attached to a non-terminal heavy atom (a non-hydrogen atom).
4	numHetero	rdMolDescriptors.CalcNumHeteroatoms(mol)	Number of heteroatoms in the molecule. Heteroatoms are atoms that are neither carbon nor hydrogen.
5	LogP	Crippen.MolLogP(mol)	Logarithm of the octanol-water partition coefficient: This metric measures a compound's lipophilicity. High LogP values indicate that a compound is hydrophobic (water-insoluble).
6	numHdon	rdMolDescriptors.CalcNumHBD(mol)	Number of hydrogen bond donors in the molecule: A hydrogen bond donor is an atom capable of donating a hydrogen atom to form a hydrogen bond.
7	numHacc	rdMolDescriptors.CalcNumHBA(mol)	Number of hydrogen bond acceptors in the molecule: A hydrogen bond acceptor is an atom that can accept a hydrogen atom to form a hydrogen bond.
8	sa	rdFreeSASA.CalcSASA(mol, radii, opts)	Solvent accessible surface area determined utilizing the Shrake-Rupley algorithm [155] , representing the portion of the molecule's surface that can be accessed by a solvent.
9	psa	rdMolDescriptors.CalcTPSA(mol)	Topological polar surface area: This

Index	Descriptor	RDKit Function	Description
			refers to the surface area occupied by polar atoms in the molecule, typically oxygen and nitrogen, along with their attached hydrogens.
10	apsa	sa - psa	Apolar surface area is determined by subtracting the polar surface area (psa) from the total solvent accessible surface area (sa).
11	psa_div_sa	psa / sa	The ratio of polar surface area to the total solvent accessible surface area provides insight into the polarity of the molecule's surface.
12	HeteroFr	numHetero / numAtom	Fraction of heteroatoms in the molecule. It gives an idea of the molecule's composition in terms of heteroatoms.

By focusing on a minimal yet effective set of input variables, including molecular descriptors derived from the chemical structures of formers, as well as essential operational parameters such as mole fraction and pressure, we aim to predict the equilibrium temperature.

The molecular weight (mw) of a molecule is determined by summing the atomic weights of all its atoms. The total number of atoms (numAtom), including hydrogens, can be counted using a function. Another important descriptor is the number of rotatable bonds (numRT), defined as single bonds that are not part of a ring and are connected to non-terminal heavy atoms. Additionally, the number of heteroatoms (numHetero), which are atoms other than carbon and hydrogen, can be calculated to assess a molecule's composition.

The LogP value measures the lipophilicity of a molecule, with higher values indicating greater hydrophobicity. Other key features include the number of hydrogen bond donors (numHdon) and acceptors (numHacc), which represent atoms that can donate or accept hydrogen atoms in hydrogen bonds, respectively. The solvent-accessible surface area (sa) is determined by calculating the portion of the molecule's surface that can be accessed by solvents, while the topological polar surface area (psa) specifically measures the area occupied by polar atoms, typically oxygen and nitrogen, along with their attached hydrogens. The apolar surface area (apsa) is derived by subtracting the psa from the total solvent-accessible surface area. The ratio of polar surface area to total surface area (psa_div_sa) provides insight into the molecule's polarity. Finally, the fraction of heteroatoms (HeteroFr), calculated as the ratio of heteroatoms to the total number of atoms, indicates the heteroatom composition within the molecule.

Table 3-3 summarizes the statistical properties of various molecular descriptors for 33 distinct chemicals. Each descriptor exhibits typical distribution characteristics, with some variations. Most descriptors follow relatively normal distributions. However, mw has higher kurtosis (2.84), indicating a more peaked distribution, and numHetero shows positive skewness (1.70) and kurtosis (3.30), suggesting a right-skewed distribution with outliers. LogP exhibits negative skewness (-0.70), while the polar surface area (psa) shows both high skewness (1.93) and kurtosis (4.54), reflecting a long right tail and a sharp peak. It is important to note that these statistics are based on 33 distinct chemicals, each with a different number of samples in the dataset. Consequently, the values for these descriptors may vary across the entire sample set.

Table 3-3: Statistical parameters of the molecular descriptors used in this study.

Descriptor	min	max	median	mean	skewness	kurtosis
mw	32.04	174.20	82.10	85.79	1.13	2.84
numAtom	6	26	14	14.47	0.62	-0.12
numRT	0	9	1	1.91	1.35	1.36
numHetero	1	6	1	1.85	1.70	3.30
LogP	-1.67	1.19	0.40	0.13	-0.70	-0.65
numHdon	0	4	1	1.06	0.90	0.51
numHacc	1	4	1	1.79	1.01	-0.17
sa	69.95	261.63	140.26	146.34	0.85	0.90
psa	3.24	127.72	20.23	30.59	1.93	4.54
apsa	49.72	184.76	113.98	115.75	0.25	-0.91
psa_div_sa	0.02	0.56	0.16	0.21	0.96	0.11
HeteroFr	0.05	0.30	0.11	0.13	0.56	-0.70

3.3.3. Data splitting

The data splitting process is crucial in the development of an ML model, as it significantly influences the model's accuracy and generalizability. A well-considered data splitting strategy ensures that the model has access to sufficient training data to avoid underfitting while also ensuring that the testing data can reliably evaluate the model's ability to generalize beyond the training set, thus preventing overfitting.

In alignment with methodologies adopted in previous studies [110,157,158] which consider an external testing set, consisting of data for chemicals not used during the training phase, a new approach was employed to rigorously evaluate the model's accuracy and, more importantly, its generalizability. This approach differs from traditional sample-based splits, focusing instead on a former-based approach. In the current study, the total collection of 34 formers was divided in an

80:20 ratio for training and testing, respectively. Consequently, data from 27 formers were used for training, while data from the remaining 7 formers were exclusively reserved for testing the model's performance. This meticulous approach to data splitting underscores the emphasis on ensuring the model's generalizability. If the model exhibits satisfactory performance on the testing set, it can be inferred that the model is likely to maintain comparable accuracy for any hypothetical former not included in the study, whose data might become available in the future. Numerically, this division resulted in 736 data samples allocated for training and 111 for testing, indicating that approximately 13% of the data samples were utilized for testing. This strategic distribution of data samples and formers is a deliberate move to enhance the robustness and applicability of the proposed model.

3.3.4. Feature transformation

In our current study, we have addressed the challenge of sample distribution in feature space, a concern identified by Pornaroontham et al. [132]. Their approach involved the exclusion of interesting outlier samples to achieve a distribution closer to normality. However, in our study, we have implemented feature transformation techniques to normalize the distribution without omitting any data. This approach not only addresses the issue of skewed distributions but also accommodates to the non-linear relationships between features and targets, improving upon the singular squared transformation employed in the aforementioned study. We have explored a range of transformations to better suit the diverse nature of our data, as summarized in Table 3-4. This table clarifies the transformations applied, their respective formulas, and the naming conventions implemented in our study.

Table 3-4: Transformations used in this study.

Transformation	Formula	Naming Convention
None	x	$x0$
Inverse	$\frac{1}{x}$	$x1$
Square Root	$\sqrt{x}$	$x2$
Square	x^2	$x3$
Natural Logarithm	$\ln(x)$	$x4$
Exponential	$\exp(x)$	$x5$

It is crucial to note that not all transformations align with the goal of normalizing data distribution. In fact, some can exacerbate the deviation from normality, as they inherently function in contrasting directions, such as the logarithmic (ln(x)) and exponential (exp(x)) transformations. Consequently, it is imperative to incorporate a method for selecting features that exhibit distributions close to normality. This selective approach ensures that the transformation process contributes positively to model accuracy and generalizability.

3.3.5. Feature filtering

In this step, we will decide which independent variables to keep, emphasizing the importance of making these decisions exclusively on the training data. Three systematic methods will be employed to refine the feature set by excluding those that are less relevant. These methods are designed to identify and remove features that (i) exhibit significant deviation from a normal distribution, (ii) demonstrate high multicollinearity, or are (iii) considered less impactful according to the feature importance rankings from a Random Forest (RF) model. This tripartite approach ensures a comprehensive and careful selection of features, thereby enhancing the model's performance and interpretability.

The initial step involves analyzing the descriptors for the 34 formers, aiming to filter out features that do not meet the established criteria. Following this, the operational parameters, specifically the mole fraction and pressure, will be evaluated using the same rigorous methodology. This careful and methodical approach ensures that only the most reliable variables are selected, thereby enhancing the robustness and predictive accuracy of the model.

The skewness and kurtosis serve as statistical indicators to assess the conformity of data to a normal distribution. Skewness measures the asymmetry or deviation from symmetry in the distribution, indicating whether the data is spread out more on one side. On the other hand, kurtosis measures the 'tailedness' of the distribution, identifying the presence of outliers that significantly differ from the majority of data points. The ideal value for both aforementioned parameters is 0. Table 3-5 lists the skewness and kurtosis for the independent variables. For instance, the feature 'psa0' initially exhibits a substantial degree of skewness and kurtosis, with values of 1.78 and 3.28, respectively. However, applying an inverse transformation moderates these values to 1.03 and

0.48, suggesting a distribution closer to normality. Conversely, an exponential transformation on the same feature amplifies the skewness and kurtosis to 4.90 and 22.04, respectively, indicating a distribution with greater deviation and heavier tails.

To reduce the influence of features with highly irregular distributions, a filtering criterion based on the combined measure of skewness and kurtosis was established. Specifically, features were selected for subsequent analysis only if the sum of their absolute values of skewness and kurtosis was less than 3. This threshold ensures the exclusion of features with extreme distributions, enhancing the robustness of the model. This value is arbitrary and selected such that variables with extreme samples are removed while maintaining an adequate number of features. Although a value closer to zero can be selected, a stricter criterion would result in losing valuable information. The features not meeting this criterion are prominently highlighted in bold in Table 3-5, providing a clear description of the data characteristics considered most suitable for the predictive model.

Table 3-5: The skewness and kurtosis for the original and transformed formers' descriptors.

Feature	Skew	Kurtosis	Feature	Skew	Kurtosis	Feature	Skew	Kurtosis
mw0	0.96	1.55	LogP0	-0.5	-1.03	**psa0**	**1.78**	**3.28**
mw1	**1.94**	**5.04**	**LogP1**	**-4.72**	**20.82**	psa1	0.46	-0.98
mw2	0.31	0.85	LogP2	-0.02	-1.25	psa2	1.03	0.48
mw3	**2.04**	**4.57**	LogP3	1.18	0.67	**psa3**	**3.33**	**11.55**
mw4	-0.41	1.12	LogP4	-0.5	-1.03	psa4	0.42	-0.84
mw5	**4.9**	**22.04**	LogP5	0.43	-1.03	**psa5**	**4.9**	**22.04**
numAtom0	0.52	-0.42	numHdon0	0.76	0.07	apsa0	0.15	-1
numAtom1	**1.22**	**2.27**	numHdon1	0.57	-1.25	apsa1	1.17	0.88
numAtom2	0.19	-0.55	numHdon2	-0.31	-1.14	apsa2	-0.13	-0.87
numAtom3	1.11	0.62	**numHdon3**	**2.24**	**5.02**	apsa3	0.59	-0.91
numAtom4	-0.2	-0.27	numHdon4	-0.07	-1.07	apsa4	-0.45	-0.54
numAtom5	**4.77**	**21.15**	**numHdon5**	**3.5**	**12.59**	**apsa5**	**4.6**	**19.97**
numRT0	1.13	0.39	numHacc0	0.83	-0.58	psa_div_sa0	0.87	-0.18
numRT1	0.24	-1.65	numHacc1	-0.27	-1.54	psa_div_sa1	0.72	-1.05
numRT2	0.23	-1.25	numHacc2	0.59	-1.03	psa_div_sa2	0.38	-0.87
numRT3	**2.42**	**6.02**	numHacc3	1.33	0.57	**psa_div_sa3**	**1.65**	**1.64**
numRT4	0.28	-1.31	numHacc4	0.52	-1.16	psa_div_sa4	-0.08	-1.19
numRT5	**4.85**	**21.7**	**numHacc5**	**1.94**	**2.42**	psa_div_sa5	1.11	0.32
numHetero0	**1.5**	**2.08**	sa0	0.71	0.21	HeteroFr0	0.47	-0.71
numHetero1	-0.36	-1.34	sa1	1	1.82	HeteroFr1	0.73	-0.82
numHetero2	0.96	0.19	sa2	0.35	-0.11	HeteroFr2	0.12	-1.14
numHetero3	**2.76**	**8.26**	**sa3**	**1.37**	**1.7**	HeteroFr3	1.27	1.16

Feature	Skew	Kurtosis	Feature	Skew	Kurtosis	Feature	Skew	Kurtosis
numHetero4	0.79	-0.35	sa4	-0.05	-0.03	HeteroFr4	-0.19	-1.26
numHetero5	**4.66**	**20.41**	**sa5**	**NaN**	**NaN**	HeteroFr5	0.59	-0.48

The subsequent phase involves eliminating features that exhibit high collinearity, utilizing the Pearson correlation coefficient (R) for this purpose. This parameter, ranging from -1 to 1, shows the degree of correlation with the direction. While 0 indicates no correlation, -1 and 1 show strong negative and positive correlations between two variables, respectively. We set a threshold at 0.95; pairs of features with an absolute R value exceeding this threshold are considered highly correlated, requiring the exclusion of one feature from each pair. Using this threshold value, we remove the features that are highly correlated. Employing a more restrictive value will lead to excluding more features and, consequently, losing the valuable knowledge they might provide for the model to learn. The decision on which feature to exclude is based on the combined measure of skewness and kurtosis, with the feature displaying the higher sum of these values being removed. Figure 3-2 illustrates the correlation matrix, showing the R values among the retained features.

In terms of feature count: initially, we identified 12 primary features as outlined in Table 3-2. Subsequent to applying five distinct feature transformations to each, the total number of features expanded to 72. The process of assessing normality led to the exclusion of 21 features, reducing the count to 51. Further refinement using collinearity analysis excluded an additional 38 features, leaving a final set of 13 features that effectively describe the formers in the database, including mw2, numRT0, numHetero4, LogP2, LogP3, numHdon0, sa4, psa1, psa4, psa_div_sa0, HeteroFr5, numRT2, and LogP4.

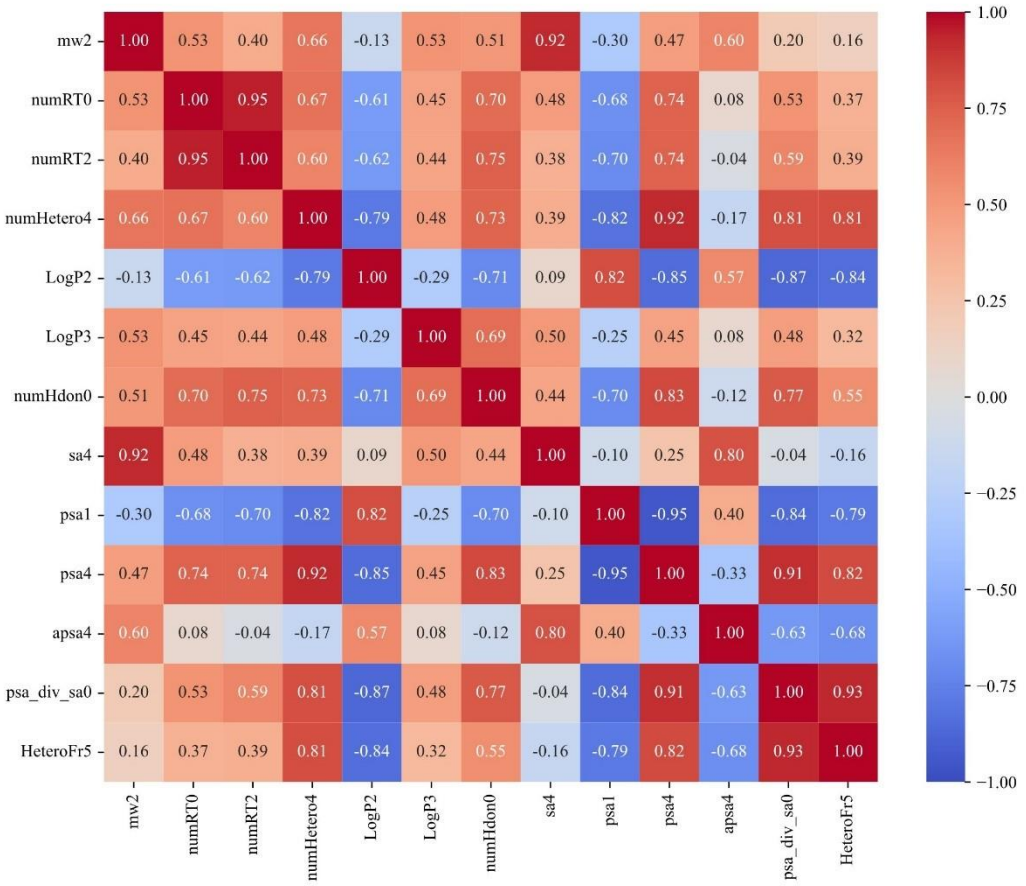


Figure 3-2: Correlation matrix of selected features.

Following the feature selection process for the descriptors, the operational parameters mole fraction and pressure underwent the same rigorous analysis. Using data from 736 training samples, both parameters were examined for their distribution characteristics. As shown in Table 3-6, mole fraction and pressure exhibited significant skewness with extreme values, which previously caused Pornaroontham et al. [132] to exclude some data points. However, after applying transformation techniques—especially the logarithmic transformation for pressure—the data distributions were effectively normalized. This allowed all data samples to be retained for modeling, maximizing data use without discarding extreme values. Further analysis of collinearity revealed a high Pearson correlation (R = 0.954) between mol_fr2 and mol_fr4, indicating strong linear dependence. Based on their distribution characteristics, mol_fr2 was identified as the less

favorable feature and excluded to reduce redundancy and avoid multicollinearity issues. To clarify, the bold formatting in Table 3-6 highlights features with extreme skewness or kurtosis before transformation, not those excluded due to collinearity. Although mol_fr2 is not bolded in the table, it was excluded from the final dataset. The bolded features are noted in the table for ease of reference.

Table 3-6: The skewness and kurtosis for the original and transformed operational parameters.

Feature	Skew	Kurtosis
mol_fr0	**2.42**	**6.88**
mol_fr1	**2.22**	**5.57**
mol_fr2	1.21	1.48
mol_fr3	**4.77**	**27.12**
mol_fr4	0.00	-0.38
mol_fr5	**2.82**	**9.71**
P_(MPa)0	**4.70**	**32.15**
P_(MPa)1	**6.77**	**65.78**
P_(MPa)2	**1.71**	**7.65**
P_(MPa)3	**9.33**	**96.13**
P_(MPa)4	-0.44	2.02
P_(MPa)5	**22.47**	**535.13**

A total of 15 features, including the operational parameters, were selected for modeling.

The feature filtering based on the RF model feature importance will be discussed in section 4.3.

3.3.6. Scaling

In the context of ML, feature scaling is a crucial preprocessing step, especially when the algorithms in use are sensitive to the scale of the input data. Distance-dependent methods such as MLP and SVM are particularly impacted by the scale of the features because these algorithms rely on calculating distances between data points. When features have diverse ranges, a model might excessively weigh the features with larger scales more heavily, leading to skewed or suboptimal results. For instance, in a dataset where one feature is measured in thousands and another in fractions, the distance calculations will be dominated by the larger-scale feature, potentially obscuring the importance or relevance of the smaller-scale feature.

Standard scaling, the method chosen for this study, addresses this issue by transforming the data such that each feature has a mean (μ) of 0 and a standard deviation (σ) of 1. The transformation is done by subtracting the mean and dividing by the standard deviation for each feature:

$$z = \frac{(x - \mu)}{\sigma} \tag{3-1}$$

where z is the standardized value, x is the original value, μ is the mean of the feature, and σ is the standard deviation of the feature.

After scaling, all features will be on a similar scale, ensuring that no single feature dominates the model due to its scale. This uniformity allows distance-dependent algorithms like MLP and SVM to operate more effectively, as the Euclidean distance between feature vectors now genuinely represents the similarity or dissimilarity between data points, without deviation from the range of the values. Moreover, this scaling can contribute to faster convergence of gradient descent during model training, as it ensures a more balanced and uniform gradient across all features.

3.3.7. SHapley Additive exPlanations (SHAP) method

The Shapley Additive Explanations (SHAP) approach is indeed used to quantify feature importance and analyze the effect of features, which are utilized as inputs, on the output of the system. Unlike traditional methods, SHAP values directly depend on input features and ML model outputs, rather than the model structure itself. This is particularly advantageous in ensemble models where specific prediction equations are unavailable. Developed from coalitional game theory, SHAP provides a systematic framework for interpreting ML model predictions, thereby offering flexibility across a wide range of ML models [159,160].

The methodology is shown in Figure 3-3.

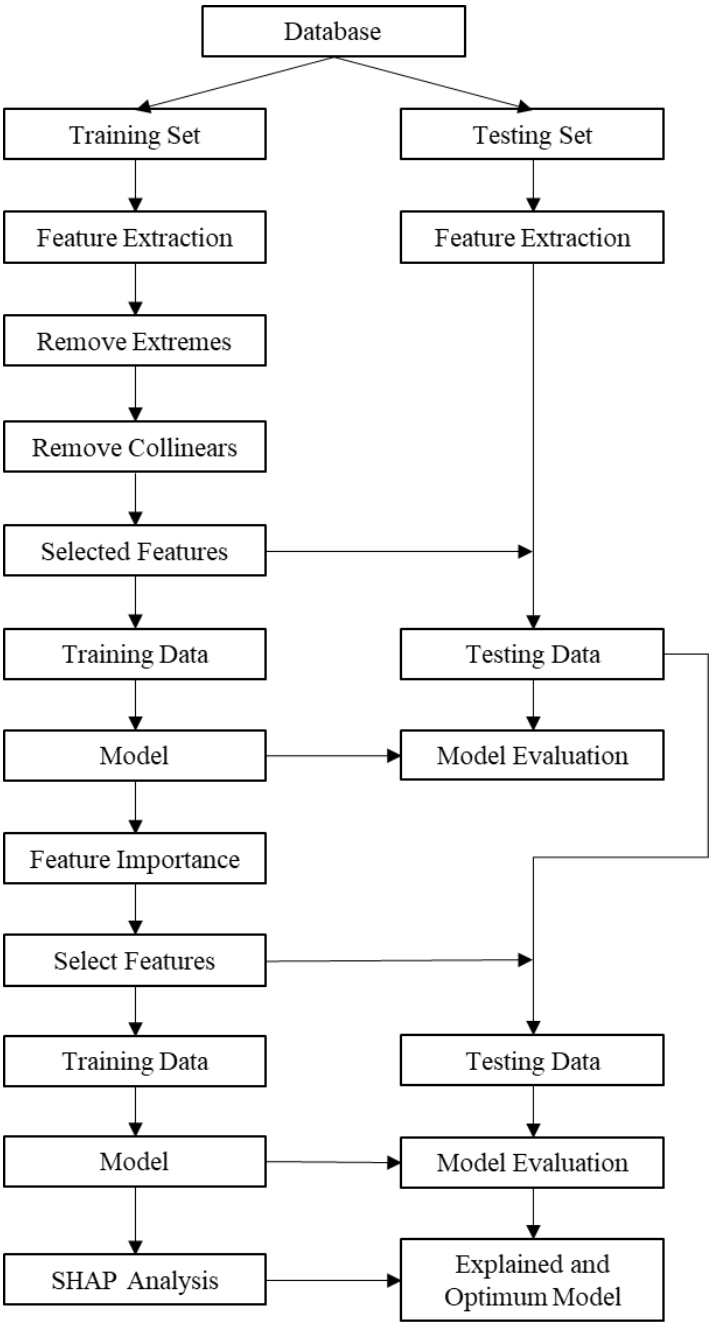


Figure 3-3: Flowchart of research methodology.

3.4. Results and discussion

3.4.1. Model development

In this section, we explore the development of ML models utilizing the Scikit-Learn library in Python [79]. Our focus encompasses four distinct predictive models: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron (MLP). The development process is methodically structured into three primary stages:

The initial stage, hyperparameter optimization, is dedicated to the meticulous tuning of crucial model hyperparameters, a process vital for enhancing model performance. By fine-tuning these hyperparameters, we aim to customize each model to the unique characteristics of our dataset, thereby optimizing their predictive abilities.

The next step is model fitting. There, we proceed with the actual training of the models. Recognizing the stochastic nature of many ML algorithms, we conduct multiple iterations of training, ensuring robustness and reliability in our models. The best-performing iteration of each model, determined through predefined criteria, is then preserved for further analysis.

The final step is accuracy analysis. In this concluding phase, the performance of the models is rigorously evaluated. Employing both graphical and statistical methods, we examine the models across two sets of training and testing data. This evaluation, along with the approach used for data splitting, allows us to assess not just the accuracy but also the generalizability of each model, ensuring that our final selections are not only precise but also robust across varied data scenarios.

3.4.1.1. Hyperparameter optimization

Hyperparameters are essential in defining a model's external structure and are set before the training phase begins, in contrast to model parameters that are learned during training. These hyperparameters significantly influence the architecture of the model, and selecting the right combination is critical for the model's optimal performance. It is crucial to note that the process of hyperparameter optimization is conducted exclusively using the training dataset to prevent model overfitting and to ensure that the model generalizes well to new, unseen data.

Table 3-7 presents a detailed summary of the selected hyperparameters along with their respective search ranges for each predictive model. We employed a randomized search strategy over 100 iterations to optimize these hyperparameters. This table notably features the best results from a 5-fold Cross-Validation using the Root Mean Squared Error (RMSE) metric, highlighting the efficacy of the models under consideration. It is essential to recognize that the error metrics reported are based on the scaled dataset.

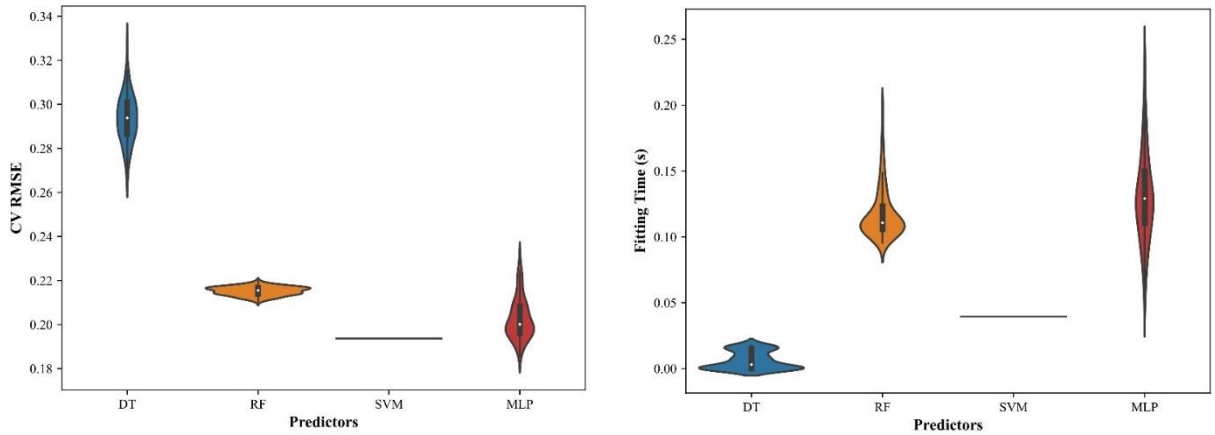
Table 3-7: The determined hyperparameters as well as the searching space for each model.

Hyperparameter	Search Interval	DT	RF	SVM	MLP
max_depth	3—20	14	13	—	—
max_features	0.01—1	0.75	0.58	—	—
n_estimators	4—100	—	76	—	—
C	0.1—1000	—	—	151.73	—
Gamma	0.1—1000	—	—	0.14	—
hidden_layer_sizes	3—30	—	—	—	25
Alpha	1e-05—1.0	—	—	—	0.0096
learning_rate_init	1e-05—0.1	—	—	—	0.0090
Best CV RMSE	—	0.273	0.212	0.194	0.197
Elapsed Time (s)	—	5.29	17.17	9.06	37.94

3.4.1.2. Model fitting

The ML models went through training with scaled input values, and Figure 3-4 presents a violin plot that visualizes the 5-fold Cross-Validation Root Mean Squared Error (CV RMSE) from 100 iterations for the DT, RF, SVM, and MLP models. This violin plot not only captures the distribution of the RMSE values but also enhances the analysis beyond what a traditional box plot offers by showing the distribution.

From the plot, the SVM and MLP models emerge as the top performers, achieving a mean 5-fold CV RMSE of about 0.194 and 0.197, respectively. They are closely followed by the RF model, which recorded mean 5-fold CV RMSEs of 0.212. The RF model's advantage is not limited to its accuracy; it also demonstrates remarkable precision, as evidenced by the tight distribution of its 5-fold CV RMSE values. As is depicted in Figure 3-4 the RMSE values for RF are much more compact than those for the MLP. This indicate the consistent results when the models are run with different initialization.



(a) (b)

Figure 3-4: The violin plot for the results acquired from 100 trials for the ML models (a) 5-fold CV RMSE and (b) fitting time.

3.4.2. Accuracy assessment

Different statistical parameters were employed to assess prediction accuracy. This study utilizes Residual (Res_i), Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), Average Relative Deviation (ARD), Average Absolute Relative Deviation (AARD), and Mean Absolute Error (MAE) to evaluate prediction performance. The formulations for these parameters are outlined below:

$$Res_i = t_i - y_i \tag{3-2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Res_i)^2}{\sum_{i=1}^{n}(\bar{t} - t_i)^2} \tag{3-3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Res_i)^2} \tag{3-4}$$

$$ARD = \frac{100}{n} \times \sum_{i=1}^{n}\frac{Res_i}{t_i} \tag{3-5}$$

$$AARD = \frac{100}{n} \times \sum_{i=1}^{n}\frac{|Res_i|}{t_i} \tag{3-6}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Res_i| \tag{3-7}$$

where y and t are the model's output and target value, respectively.

Table 3-8 shows statistical metrics that evaluate the accuracy of the models developed in this study. These metrics are calculated based on the original values, not the scaled data used during the training phase. The purpose of this comparative analysis is to evaluate the models' performance on the testing dataset, with the best values distinctly highlighted in bold for emphasis. In this assessment, the RF model stands out for its lowest RMSE at 1.71 K, showing consistent superiority across all considered parameters. This achievement is particularly notable when compared to the performances of the SVM and MLP models, which had higher error values of 5.74 K and 4.34 K, respectively.

It is crucial to acknowledge that the CV RMSE values, discussed earlier, were exclusively based on the training dataset. This implies that the data associated to the formers in the validation set had previously been encountered by the model during training. However, the real test of robustness comes from the testing set, where all formers are entirely novel to the model. Here, while the RF model shows slightly better performance than the DT, both models demonstrate relatively strong generalization to unseen formers compared to SVM and MLP. This highlights the importance of appropriate data splitting in validating model effectiveness and ensuring reliable performance on novel data. The observed differences in metrics underscore how the choice of validation strategy can impact a model’s ability to generalize.

Table 3-8: The statistical parameters showing the accuracy of the models developed in this study.

Model	Set	AARD (%)	ARD (%)	RMSE	MAE	R^2	N
DT	Training	0.01	0.00	0.28	0.02	0.999	736
	Testing	0.53	-0.48	1.93	1.52	0.911	111
RF	Training	0.16	0.03	0.78	0.45	0.993	736
	Testing	**0.48**	-0.39	**1.71**	**1.37**	**0.930**	111
SVM	Training	0.29	0.01	1.46	0.82	0.977	736
	Testing	1.17	-1.07	5.74	3.42	0.209	111
MLP	Training	0.28	0.03	1.51	0.81	0.975	736
	Testing	0.68	**0.05**	4.34	1.96	0.547	111

RF model's outstanding performance is due to its inherent characteristics, particularly its ensemble approach, which is proficient at capturing complex data relationships, leading to improved predictive accuracy compared to other models in this study.

3.4.3. Reduced model

Previously, it was mentioned that another feature selection process, guided by the feature importance metrics derived from the Random Forest (RF) model, should be conducted. Figure 3-5, illustrating these metrics, shows that certain features have minimal impact on the final outcomes, indicating that they could be excluded from the dataset. It is possible to select influential features based on Figure 3-5. This plot shows that after the 7th feature, the features' contributions drop significantly. Among these 7 features, the first 4 are comparatively more important. As we aim to select as few features as possible, and the first 4 influential ones cover all operational features as well as descriptors, by setting a threshold of 0.12 for feature importance, we keep four significant features for further modeling.

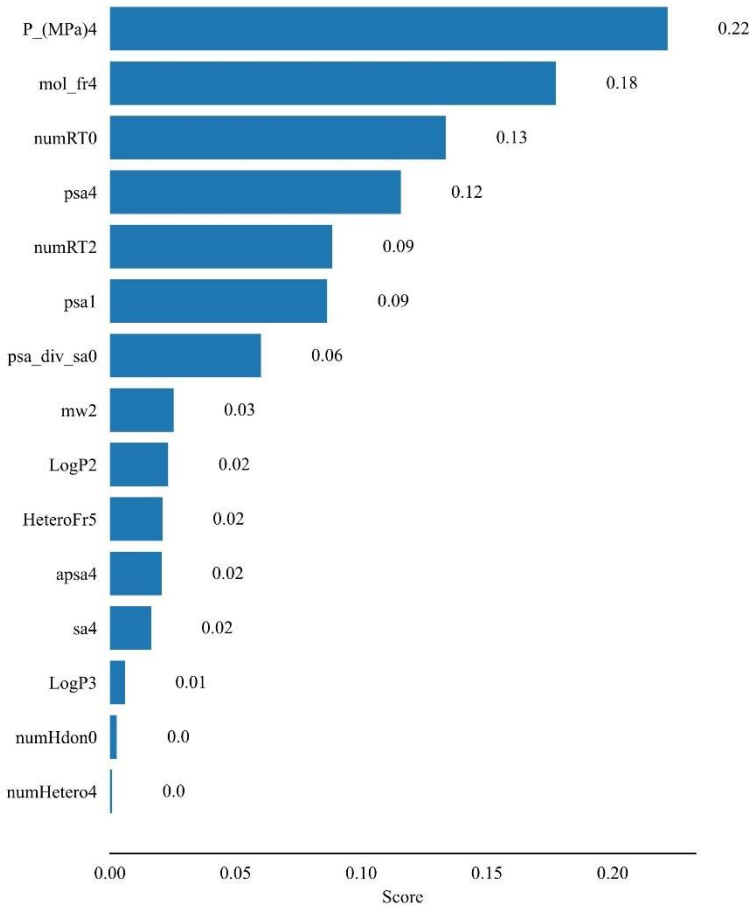


Figure 3-5: Feature importance metrics based on the developed RF model.

We carefully followed the model development protocol outlined in the earlier sections, utilizing the refined feature set. Table 3-9 lists the optimized hyperparameters for each model type. Following the determination of the optimal model configurations, each model underwent 100 training iterations with varied initializations. The SVM and MLP models showed the best performance in this refined setup, resulting the most favorable CV RMSE values of 0.23 and 0.24, respectively, while the RF model closely followed with a CV RMSE of 0.25.

Table 3-9: The determined hyperparameters as well as the searching space for each reduced model.

Hyperparameter	Search Interval	DT	RF	SVM	MLP
max_depth	3—20	9	10	—	—
max_features	0.01—1	1.00	0.75	—	—
n_estimators	4—100	—	57	—	—
C	0.1—1000	—	—	79.98	—
gamma	0.1—1000	—	—	0.32	—
hidden_layer_sizes	3—30	—	—	—	25
alpha	1e-05—1.0	—	—	—	0.01
learning_rate_init	1e-05—0.1	—	—	—	0.01
Best CV RMSE	—	0.32	0.25	0.23	0.24
Elapsed Time (s)	—	6.54	19.09	12.85	50.25

The evaluation of model performance did not end there; we further compared the testing RMSE of the original and reduced models, as presented in Table 3-10. This comparison reveals that the error rates for the DT and RF models display minimal difference between the original and reduced versions. However, there is a significant increase in error rates for the distance-based MLP and SVM models in their reduced forms. Figure 3-6 schematically presents the comparison of AARD, RMSE, and MAE. Considering the MAE, the errors in all the reduced models have increased. The comparison of AARD, however, shows a different trend. While the AARD values for RF and MLP have decreased, there is an increasing trend for DT and SVM. All these three metrics indicate that the reduced RF model exhibits a comparative error similar to that of the original RF model.

Table 3-10: Comparison of the testing RMSE between the original and reduced models.

Model	Testing RMSE	
	Original Model	Reduced Model
DT	1.93	2.40
RF	1.71	1.63
SVM	5.74	10.14
MLP	4.34	12.98

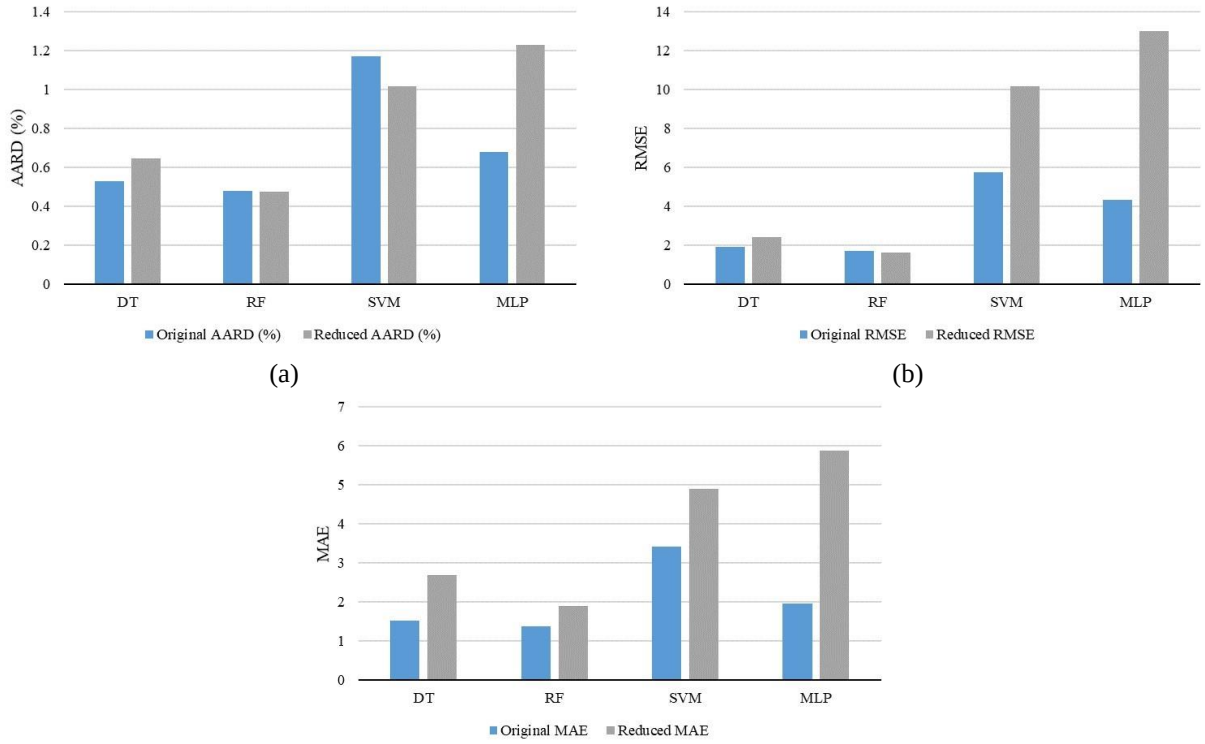


Figure 3-6: Systematic comparison of the original and reduced models based on (a) AARD (%), (b) RMSE, and (c) MAE.

The assessment of model performance extends beyond the training phase, involving a meticulous evaluation of how each model, both in its original and reduced forms, performs on unseen data. Table 3-10 is crucial here, as it compares the testing RMSE for the original models with their reduced versions, providing a clear perspective on how feature reduction affects model effectiveness. From Table 3-10, it is evident that the DT and RF models exhibit consistent performance, with minimal difference in error rates between the original and reduced models. This

consistency indicates a robust feature set where even a reduced number of features can capture the essential patterns in the data. However, a notable observation stands out regarding the MLP and SVM models. The considerable increase in error rates when using reduced feature sets indicates a significant sensitivity to the dimensionality of the feature set. This rise in error rates for the distance-based models may suggest that the removed features, despite being considered less important according to the RF model's feature importance metrics, contained valuable information essential for the performance of MLP and SVM models. This observation highlights the intricate balance between feature relevance and model dependency in MLP and SVM models.

Moreover, it is crucial to consider the role of data splitting, as discussed earlier. The CV RMSE values, derived from the training dataset, offer an initial understanding of model performance. However, the ultimate proof to a model's robustness and reliability is its performance on the testing set, comprising entirely novel formers. This phase truly assesses the model's generalizability and its ability to extrapolate its learned patterns to new, unseen scenarios. As demonstrated by the RF model's performance, meticulous data splitting and feature selection are essential to ensure that a model not only learns effectively but also reliably generalizes its insights to new data, maintaining accuracy and precision in its predictions.

3.4.4. Model complexity reduction by feature selection

The SHAP method helps us identify which factors tend to influence the equilibrium towards higher temperatures. SHAP values are available for each feature and each sample. A negative SHAP value for an instance indicates that the feature has a negative impact on the target value, in this study, the hydrate equilibrium temperature. Conversely, a positive SHAP value signifies a positive impact of the feature on the target value for that instance.

The SHAP analysis was performed on the optimum reduced model. The provided SHAP beeswarm plot in Figure 3-7 shows the impact of different input features on the model predicting the hydrate equilibrium temperature. Since the data values are standardized, the SHAP values indicate how much higher or lower the standardized feature value would shift the predicted equilibrium temperature from the base value (which is the average prediction of the model for the training data).

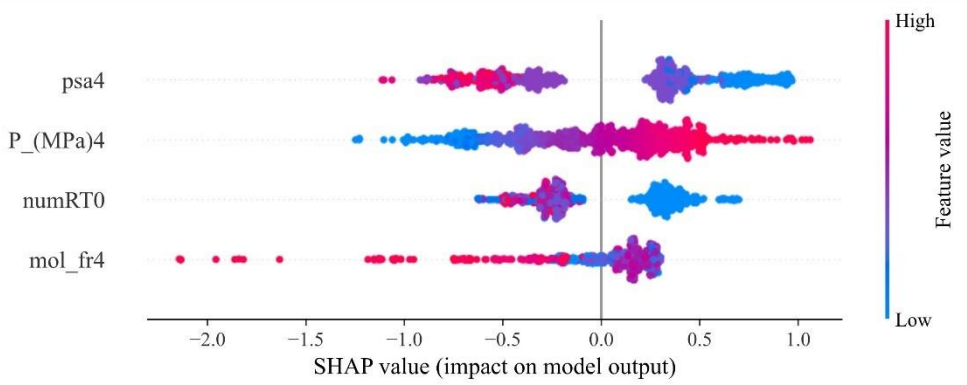


Figure 3-7: Beeswarm plot visualizing the distribution of SHAP values for key predictive features.

The SHAP dependence plot, depicted in Figure 3-8, can provide more detailed information about the trends. The SHAP values for pressure increase almost monotonically as the logarithm of pressure increases, which suggests a logarithmic relationship between pressure and hydrate equilibrium temperature. Physically, this could mean that the hydrate formation is highly sensitive to changes in pressure at lower values, with this sensitivity decreasing at higher pressures due to the logarithmic transformation, reflecting a typical gas solubility behavior in liquids.

The plot indicates that as the standardized mol_fr4 value increases, the SHAP values tend to decrease. The decrease in SHAP value signifies a negative correlation with the hydrate equilibrium temperature; as the mol_fr4 increases, the predicted equilibrium temperature tends to be lower. There is a significant concentration of high-negative SHAP values at higher levels of mol_fr4, indicating a strong influence of high mol_fr4 values in reducing the predicted temperature. From a physical chemistry standpoint, mol_fr4 appears to be related to the mole fraction of an additive following a logarithmic transformation. If the additive is an inhibitor of hydrate formation, its increased presence (as indicated by a higher mole fraction) would logically suppress the hydrate equilibrium temperature. This suppression could be due to the inhibitor interfering with the water molecules' ability to form the structured lattice that constitutes the hydrate, thus stabilizing the liquid phase over the solid hydrate at lower temperatures.

The numRTO plot depicts several SHAP values, predominantly on the positive side for lower numRTO values. This suggests that having fewer rotatable bonds tends to increase the model's prediction for the hydrate equilibrium temperature. It means, former molecules with fewer

rotatable bonds appear to influence the equilibrium towards higher temperatures. In physical terms, molecules with fewer rotatable bonds are more rigid. If the numRTO feature represents the rigidity of additives that might form clathrates or otherwise interact with hydrate formation, then this result could indicate that more rigid molecules favor conditions that lead to higher hydrate equilibrium temperatures.

The SHAP plot for `psa4`, which denotes the logarithmically transformed total polar surface area, reveals a complex, non-linear relationship with the predicted hydrate equilibrium temperature. Notably, lower `psa4` values exhibit a mix of high positive and negative impacts, indicating that minor variations in the polar surface area can significantly influence the temperature at which hydrates are stable, likely due to the molecule's capacity for polar interactions. As the `psa4` value increases, the magnitude of its effect on the model's predictions decreases, suggesting a threshold beyond which additional polar surface area does not substantially alter hydrate stability, possibly due to molecular size limitations within the hydrate structure. This pattern suggests that the polar surface area plays a critical role in hydrate formation, with its effect being most significant when the area is small and capable of engaging or disrupting the formation of hydrate cages.

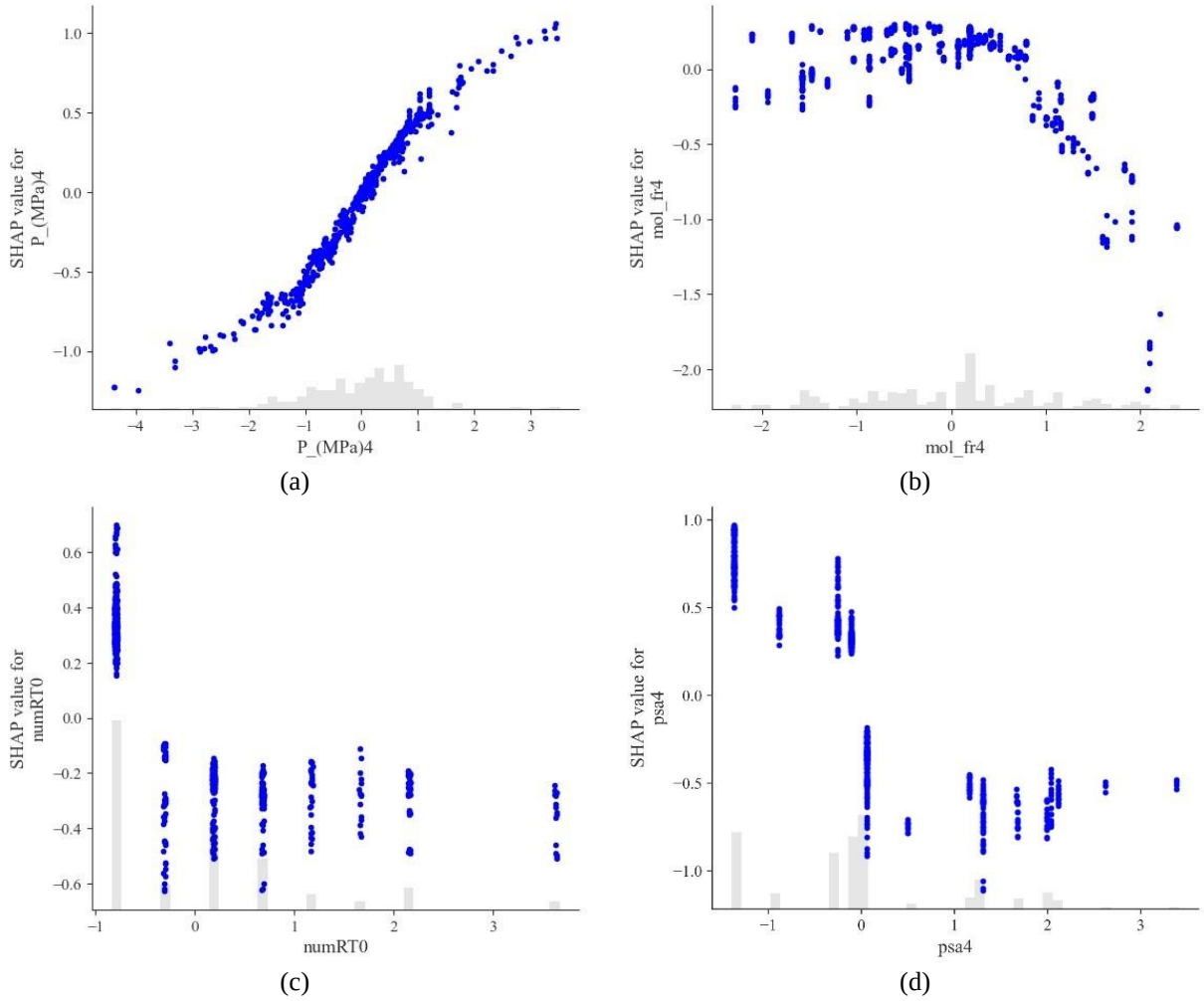
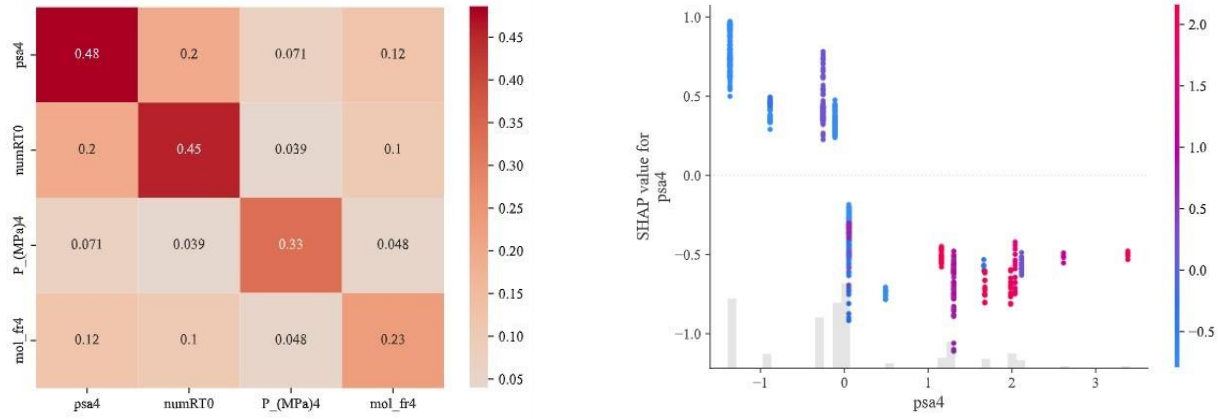


Figure 3-8: SHAP value dependency plots for features affecting hydrate equilibrium temperature prediction.

Figure 3-9 (a) displays the interaction heatmap matrix for the independent variables. In this heatmap, the main diagonal represents the primary effect of each feature, while the off-diagonal elements depict the interaction effects between different features. The colors indicate the strength of these effects. Notably, the strongest interaction is observed between psa4 and numRT0, with an interaction effect approximately half the value of their individual main effects, indicating a significant but not dominant interaction compared to their primary effects. Figure 3-9 (b) illustrates the SHAP values for psa4, plotted against the psa4 values themselves, with points colored based on numRT0 values. The color gradient provides an additional layer of insight, showing how numRT0 influences the relationship between psa4 and the target variable. It is evident that lower psa4 values are associated with lower numRT0 values, represented by the blue color. This visual

relationship highlights the interplay between psa4 and numRT0, where lower values of both variables tend to co-occur, thereby impacting the model's predictions.



(a) (b)

Figure 3-9: (a) Interaction heatmap matrix for variables, highlighting strength of effects. (b) SHAP values for psa4 colored by numRT0 levels.

This research presents a comprehensive machine learning framework for predicting the equilibrium conditions of methane hydrate systems. By accurately modeling these conditions, the study offers valuable insights for various industries. The findings related to molecular descriptors and operational parameters can inform the design of more efficient industrial processes, such as improvements in gas separation techniques, Solid Natural Gas (SNG) technology, and seawater desalination. Additionally, by identifying the most effective hydrate formers and their impact on equilibrium conditions, industries can make informed decisions about chemical additives in hydrate management, leading to more efficient and targeted applications.

3.5. Conclusions

This research focuses on developing machine learning models to accurately predict the equilibrium conditions of CH_4 clathrate hydrates, incorporating a broad range of water-soluble hydrate formers while minimizing the number of input variables. By integrating molecular descriptors derived from the chemical structures of various formers with key operational parameters such as mole fraction and pressure, we applied a diverse set of machine learning models—including Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Multi-Layer Perceptron

(MLP). This approach enabled the identification of the most statistically significant molecular descriptors for predicting hydrate dissociation temperatures with minimal deviation from experimental values. Our findings are summarized as follows:

- The study utilizes a dataset of 847 samples from 34 different hydrate formers, sourced from the literature. A novel former-based data-splitting approach was applied, allocating 27 formers for training and 7 exclusively for testing in an 80:20 ratio. This strategy ensures robust model performance and enhances generalizability across different hydrate systems.
- The study develops accurate machine learning models to predict the equilibrium conditions for CH_4 hydrates, focusing on water-soluble hydrate formers. The RF model achieved high accuracy with an R^2 of 0.930, RMSE of 1.71, and AARD of 0.48%.
- A key innovation is the use of a former-based data-splitting method, ensuring generalizability across different hydrate formers. This strategy enhances model robustness compared to traditional sample-based splits.
- By integrating feature selection and Shapley Additive Explanations (SHAP), we identified the most statistically significant molecular descriptors influencing hydrate dissociation temperatures, providing deeper insights into variable importance.
- Exploration of a reduced model highlights the importance of feature selection, particularly for distance-based models such as SVM and MLP, where less critical features still play a vital role in model functionality.
- This research advances methane hydrate technology by offering a more interpretable and accurate framework for predicting hydrate equilibrium conditions, paving the way for improved former selection and operational efficiency in gas storage and energy applications.

- Expanding the dataset to cover other gases, such as CO_2, and additional hydrate formers, along with considering environmental factors like salinity, could improve the model's generalizability and increase its relevance for industrial applications.

Chapter 4: Gas hydrate dissociation temperature prediction in porous media: effects of pore size and modeling approach

4.1. Introduction

Gas hydrates, also known as clathrate hydrates, are solid crystalline substances formed when water molecules, referred to as host molecules, arrange themselves into a cage-like structure through hydrogen bonding, effectively trapping gas or volatile liquid molecules known as guest molecules. The formation of these cages is crucial, as it stabilizes the guest molecules within the structure. Moreover, depending on the size of the guest molecules, gas hydrates can crystallize into one of three main structures: structure I (sI), structure II (sII), or structure H (sH). Each structure differs in the size and arrangement of the cages, which influences how and where these hydrates occur in nature [2].

Beyond their structural interest, gas hydrates are gaining attention for both their challenges and their potential benefits across various industries. While gas hydrates can lead to blockages in natural gas pipelines and valves [161], they also offer a wide range of industrial advantages. Their unique molecular arrangement not only enables efficient natural gas storage and transportation [162,163] but also serves as a robust method for separating complex gas mixtures through hydrate formation [164,165]. Researchers are actively exploring their use for CO_2 capture and sequestration, a promising approach to mitigate greenhouse gas emissions [166,167]. Additionally, gas hydrates are applied in water desalination processes to provide fresh water [168,169]. There is growing interest in hydrogen hydrates as a viable medium for producing green fuel, which could contribute significantly to sustainable energy solutions [170,171]. In the food industry, they help concentrate liquid solutions, enhancing product quality and processing efficiency [172,173]. Moreover, their inherent capacity to absorb and release heat makes them ideal for cool storage and air conditioning applications [174,175]. Ultimately, the most significant potential of gas hydrates lies in energy supply, where they could become a key player in meeting future energy demands [176].

Given their practical relevance, it is important to understand the environmental and geological contexts in which gas hydrates form. Natural gas hydrate deposits require low temperatures and

high-pressure conditions and are naturally found in both permafrost and marine environments [177]. Rapid warming in the Arctic-Boreal region raises concerns that methane (CH_4) released from carbon-rich soils and subsea permafrost could further accelerate climate change. This has led to significant theoretical and practical interest, particularly in the Arctic shelf region, which holds vast hydrate deposits [178,179]. Gas hydrate occurrences have been observed in muddy sediments and sand-rich, coarse-grained layers, which serve as porous media [180,181]. These porous formations not only facilitate hydrate formation but also influence CH_4 migration once released. While hydrate equilibrium in the larger pores of coarse sands mirrors that of the bulk phase, extracting hydrates from fine sands or clay requires a closer look at how pore characteristics and inhibitors work together to promote dissociation. This is crucial because the smaller pore sizes in these materials can alter heat and mass transfer properties, potentially affecting both the stability of the hydrates and the efficiency of any inhibitors used [182].

To better understand these confinement effects, numerous studies have experimentally determined the equilibrium conditions for simple clathrate hydrates across various porous media [32–38]. These findings underscore the importance of experimentally investigating the dissociation behavior of hydrates in such environments, as understanding these conditions is vital for optimizing both extraction and stability management in practical applications. For instance, Kang et al. conducted experiments to determine the phase equilibria of binary systems, CH_4 + water and CO_2 + water, in porous silica gels with nominal pore sizes of 6, 30, and 100 nm. [32]. Seo et al. measured the three-phase equilibria, consisting of hydrate (H), water-rich liquid (LW), and vapor (V), for C_2H_6 and C_3H_8 hydrates in silica gels with pore sizes of 6, 15, 30, and 100 nm to examine the effect of geometrical constraints on gas hydrate phase equilibria. The results showed that, at specified temperatures, the hydrate stability region shifts to a higher pressure range compared to bulk hydrates, with the shift varying according to pore size [38].

To complement experimental findings, various models have been developed to predict hydrate equilibrium in porous media. Clarke et al. [39] developed a model based on the van der Waals–Platteeuw (vdW-P) solid solution theory that incorporates capillary pressure effects in porous media. Their approach factors in pore radius, surface energy per unit area, and wetting angle to more accurately predict the incipient conditions for hydrate formation. Sun and Duan [183] extended the van der Waals–Platteeuw model by incorporating capillary force and salinity effects

on the chemical potentials of CH_4 and H_2O. They used the Gibbs–Thomson equation, with appropriate parameters for the hydrate–water interface, to account for capillary effects in porous sediments and applied the Pitzer model to calculate activity coefficients in the CH_4–seawater system. Li et al. [184] used two thermodynamic methods, one based on equal fugacities and the other on equal activities, to predict gas hydrate dissociation in porous media. They modeled the non-hydrate phase with the Trebble–Bishnoi equation (fugacity) or the Soave–Redlich–Kwong equation (activity), and modeled the hydrate phase by combining the van der Waals–Platteeuw model with Llamedo et al.'s capillary model. Their predictions matched experimental data well, with the fugacity method performing slightly better. Azimi et al. [181] introduced a thermodynamic approach that relies on the dissociation enthalpy of hydrates and the water activity in porous media to predict the dissociation temperature of CH_4, ethane, propane and carbon dioxide gas hydrates . Their model accounts for capillary pressure, the specific properties of the porous medium, and geometric constraints when calculating water activity. Mu et al. [182] proposed two thermodynamic models to predict hydrate equilibrium conditions, based on the Chen–Guo framework with the Hu–Lee–Sum correction. In the first model, variations in water molar volume and the temperature effects on interfacial tension are taken into account, while the second model assumes that water activity is independent of temperature. Roodbari et al. [185] employed two different thermodynamic models, namely the fugacity and activity models, based on the van der Waals–Platteeuw equations and their corresponding parameters. Their study focused on assessing how pore size influences the thermodynamic modeling of natural gas hydrate formation in porous media.

In parallel with thermodynamic efforts, the rise of machine learning (ML) has made artificial intelligence algorithms a key component in hydrate research, particularly for developing predictive models of hydrate phase equilibrium. These computational methods are not only highly practical but also crucial for addressing the complex challenges found in hydrate systems [31]. Mok et al. [186] used 13 ML models to predict the phase equilibrium of CO_2 hydrates in systems containing salts and organic inhibitors. Their work demonstrates that ML models are effective in forecasting the phase equilibrium of CO_2 hydrates, whether in the presence of individual or mixed thermodynamic inhibitors. Wu et al. [187] focused on developing mathematical models to predict hydrate formation temperature in the presence of amino acids, exploring the relationship between amino acid structural properties and their hydrate inhibition effectiveness, and identifying the most

suitable type and concentration for use. The accuracy of hydrate formation temperature predictions was evaluated using multiple linear regression (MLR) and various ML techniques, including random forest (RF), M5 Rule (M5R), and support vector machine (SVM).

Focusing more specifically on hydrate formation in porous systems, Momeni et al. [188] developed two feed-forward ANN methods to predict the equilibrium temperatures of CH_4 and ethane hydrates based on pressure and pore diameter, achieving acceptable accuracy. In the following, Mohammadi et al. [189] developed two distinct approaches for predicting gas hydrate phase equilibria in systems composed of carbon dioxide, CH_4, or ethane with pure water, under various porous media conditions (including silica gel, mesoporous silica, and porous silica glass). First, they proposed a correlation based on ice point depression (IPD) to estimate the hydrate suppression temperature due to pore effects. Second, they employed the least squares support vector machine (LSSVM) algorithm to create mathematical models for determining the dissociation pressures of these systems. Their findings revealed that while the IPD-based correlation is reliable for porous silica glass, the LSSVM models provide a more general predictive capability across all investigated systems. More recently, Wan et al. [190] investigated the phase stability of CH_4 and carbon dioxide (CO_2) hydrates under confinement using ML techniques. They built three models, including support vector machine, random forest, and gradient boosting decision tree, to predict the phase stability of these confined hydrates. Their results indicated that the support vector machine achieved the highest prediction accuracy, while the random forest performed the poorest. Consequently, they selected the support vector machine model with a 70% training set fraction, which attained over 90% accuracy in predicting the unknown phase stability of both CH_4 and CO_2 hydrates. This approach offers a quick and accurate method for determining hydrate phase stability in confined conditions for future applications.

Taken together, these studies highlight the importance of understanding dissociation behavior in confined environments. In such settings, gas hydrates behave differently from their bulk counterparts, mainly due to unique small-scale interactions that complicate the evaluation of their equilibrium dissociation conditions. Therefore, in this study, we developed two distinct approaches: one based on thermodynamic modeling using the activity model, and the other employing ML techniques to investigate CH_4 and propane (C_3H_8) hydrate dissociation conditions

in confined spaces. Additionally, SHapley Additive exPlanations (SHAP) analysis was applied to interpret the ML predictions.

4.2. Models

4.2.1. Activity model

In this method, Wilder et al. [191] proposed an activity-based model that incorporates the difference in the curvature of the solid–liquid interface during hydrate dissociation to predict hydrate equilibrium dissociation conditions in porous media. This method relies on the principle that the chemical potential of water must be the same in both the hydrate phase and the liquid phase at equilibrium.

Mathematically, at equilibrium:

$$\mu_W^H = \mu_W^L \tag{4-1}$$

where μ_W^H and μ_W^L represent the chemical potentials of water in the hydrate and liquid phases, respectively.

To rewrite this in terms of the empty hydrate lattice, we substitute the chemical potential of water in the empty lattice into the above equation and factor in the capillary pressure arising during dissociation. This leads to the following relationship [191]:

$$(\Delta\mu_w^{MT-H})_{pore} = (\Delta\mu_w^{MT-L})_{pore} \tag{4-2}$$

where $\Delta\mu_w^{MT-H} = \mu_w^{MT} - \mu_w^H$ represents the difference in water's chemical potential between the empty hydrate lattice (μ_w^{MT}) and the hydrate phase (μ_w^H), while $\Delta\mu_w^{MT-L} = \mu_w^{MT} - \mu_w^L$ is the difference between the empty lattice and the liquid phase (μ_w^L). Both differences are evaluated within the pore space (subscript "pore"), where capillary effects come into play.

Next, we express $(\Delta\mu_w^{MT-H})_{pore}$ as follows:

$$(\Delta\mu_w^{MT-H})_{pore} = (\mu_w^{MT})_{pore} - (\mu_w^{H})_{pore} = RT \sum_{m=1}^{\text{no. of cavities}} v_m \ln\left(1 + \sum_{j=1}^{\text{no. of components}} C_{mj} f_j\right) \quad (4\text{-}3)$$

where v_m representing the ratio of cavity type m to the number of water molecules in a hydrate unit. f_j is the fugacity of component j, "no. of components" is the number of hydrate-forming components, and C_{mj} is the Langmuir constant for cavity type m and guest component j. These Langmuir constants C_{mj} capture the interactions between the guest molecules and water. For temperatures ranging from 260 K to 300 K, they can be determined via the empirical correlation proposed by Parrish and Prausnitz [66].

$$C_{mj} = \frac{A}{T}\exp\left(\frac{B}{T}\right) \quad (4\text{-}4)$$

where A and B are constants. The values specific to CH_4 are provided in Table 4-1.

Table 4-1: Langmuir Constant Parameters for CH_4 and C_3H_8 in Equation 4-4 [66]**.**

Gas	Parameters	Small	Large
CH_4	A (K/atm)	3.7237×10^{-3}	1.8372×10^{-2}
	B (K)	2.7088×10^{3}	2.7379×10^{3}
C_3H_8	A (K/atm)	0.0	1.2353×10^{-2}
	B (K)	0.0	4.4061×10^{3}

The difference $(\Delta\mu_w^{MT-L})_{pore}$ is given by:

$$(\Delta\mu_w^{MT-L})_{pore} = (\Delta\mu_w^{MT-L})_{bulk} + v_l\frac{\gamma_{hl}cos\theta}{r} \quad (4\text{-}5)$$

where $(\Delta\mu_w^{MT-L})_{bulk}$ is the corresponding difference in the bulk phase, v_l is water molar volume, γ_{hl} is the hydrate–liquid interfacial tension, θ is the contact angle (often taken to be zero), and r is the pore radius.

The molar volume of pure water, v_l, in m^3mol^{-1} is determined using the following empirical equation [192]:

$$\ln v_l = -10.9241 + 2.5\times10^{-4}(T - 273.15) - 3.523\times10^{-4}(P - 0.101325) + 1.559\times10^{-7}(P - 0.101325)^2 \quad (4\text{-}6)$$

where P is the pressure in MPa, and T is the system temperature in K.

Due to the lack of a direct method for measuring surface tension in the confined spaces of porous media and the limited understanding of its exact influence, researchers have relied on various estimation approaches and assumed surface tension values when modeling gas hydrate behavior in such environments [185]. Uchida et al. [34] estimated the interfacial tension between CH_4 hydrate and water under confined conditions to be approximately 39 mJ/m^2, while Anderson et al. [35] reported a lower value of (32±3) mJ/m^2 for the same interface. In addition, Uchida et al. [37] determined the interfacial tension between C_3H_8 hydrate and water under confined conditions to be approximately 25 mJ/m^2. Chen et al. [193] proposed that surface tension (σ_{HW}) should be expressed as a function of both pore size and temperature to capture the variations in surface tension due to changes in the pore radius (r) and temperature. Their proposed relationship is given by:

$$\sigma_{HW} = \frac{a + bT}{1 + \frac{\delta}{r}} \tag{4-7}$$

where δ represents the Tolman length (0.4186 nm at 273.15 K), and a and b emprical constants. The specific values of these constants for different gases are provided in reference [193].

The term $(\Delta\mu_w^{MT-L})_{bulk}$ in Equation 4-5 can be written as:

$$\frac{(\Delta\mu_w^{MT-L})_{bulk}}{RT} = \frac{\Delta\mu_w^0}{RT_0} - \int_{T_0}^{T} \frac{\Delta h_w^{MT-L^0}}{RT^2} dT + \int_{P_0}^{P} \frac{\Delta v_w^{MT-L^0}}{RT} dP - \ln(\gamma_w x_w) \tag{4-8}$$

where $\Delta\mu_w^0$ is the reference chemical potential difference at the reference temperature T_0 and pressure P_0; R is the universal gas constant; $\Delta h_w^{MT-L^0}$ and $\Delta v_w^{MT-L^0}$ are the enthalpy and volume differences between the empty lattice and the liquid phase; γ_w is the activity coefficient of water, typically assumed to be unity because hydrate formers have low solubility in the aqueous phase; and x_w is the mole fraction of water in the liquid phase.

Furthermore, the temperature dependence of the enthalpy difference is described by [66]:

$$\Delta h_w^{MT-L^0} = \Delta h_w^0 + \int_{T_0}^{T} [-38.12 + 0.141(\mathrm{T} - 273.15)] dT \tag{3-9}$$

where Δh_w^0 is the enthalpy difference between the empty hydrate lattice and the liquid water at the reference conditions. The values for the thermodynamic properties are provided in Table 4-2.

Table 4-2: Thermodynamic characteristics at reference conditions for gas hydrates [2].

Variables	Structure I (sI)	Structure II (sII)
$\Delta \boldsymbol{\mu_w^0}$ $(\boldsymbol{J/mol})$	1263.6	882.8
$\Delta \boldsymbol{v_w^{MT-L^0}}$ $(\boldsymbol{cc/mol})$	4.6	5.0
$\Delta \boldsymbol{h_w^0}$ $(\boldsymbol{J/mol})$	–4858.9	–5202.2
$\boldsymbol{v_m}$	Small cavities: 1/23	Small cavities: 16/136
	Large cavities: 3/23	Large cavities: 8/136

Substituting Equations 4-8 and 4-5 into Equation 4-3 yields:

$$\frac{\Delta\mu_w^0}{RT_0} - \int_{T_0}^{T} \frac{\Delta h_w^{MT-L^0}}{RT^2} dT + \int_{P_0}^{P} \frac{\Delta v_w^{MT-L^0}}{RT} dP - \ln(\gamma_w x_w) - \sum_{m=1}^{\text{no. of cavities}} v_m \ln\left(1 + \sum_{j=1}^{\text{no. of components}} C_{mj} f_j\right) + v_l \frac{\gamma_{hl} cos\theta}{rRT} = 0 \quad (4\text{-}10)$$

The fugacity of the gas is determined by the Soave–Redlich–Kwong equation of state (SRK EOS) [194], and the water mole fraction in the liquid phase, x_w is found via the gas solubility:

$$x_w = 1 - x_g = 1 - \frac{f}{h_w \exp(\frac{P\bar{V}}{RT})} \quad (4\text{-}11)$$

where x_g is the gas mole fraction in the water-rich phase, f is the gas fugacity, h_w is the Henry's constant for the gas in water, and $\bar{V}$ is the partial molar volume of the gas in water.

The Henry's constant h_w can be expressed as [195]:

$$-\ln\left(\frac{H_i}{101325}\right) = H_i^{(1)} + \frac{H_i^{(2)}}{T} + H_i^{(3)} \ln T + H_i^{(4)} \quad (4\text{-}12)$$

where $H_i^{(1)}$, $H_i^{(2)}$, $H_i^{(3)}$, and $H_i^{(4)}$ are empirical constants, provided in Table 4-3**.** This equation incorporates the temperature dependence of gas solubility in water.

Table 4-3: Henry's law constants [195]**.**

Gas	$H_i^{(1)}$ (Pa)	$H_i^{(2)}$ (Pa.K)	$H_i^{(3)}$ (Pa)	$H_i^{(4)}$ (Pa.K^{-1})
CH_4	–183.7860	9112.582	25.0405	–0.00015
C_3H_8	316.4900	15922.700	44.3285	0.00000

Figure 4-1 presents the modeling workflow for predicting clathrate hydrate dissociation pressure of clathrate hydrates in porous media at a specified temperature, based on the activity model.

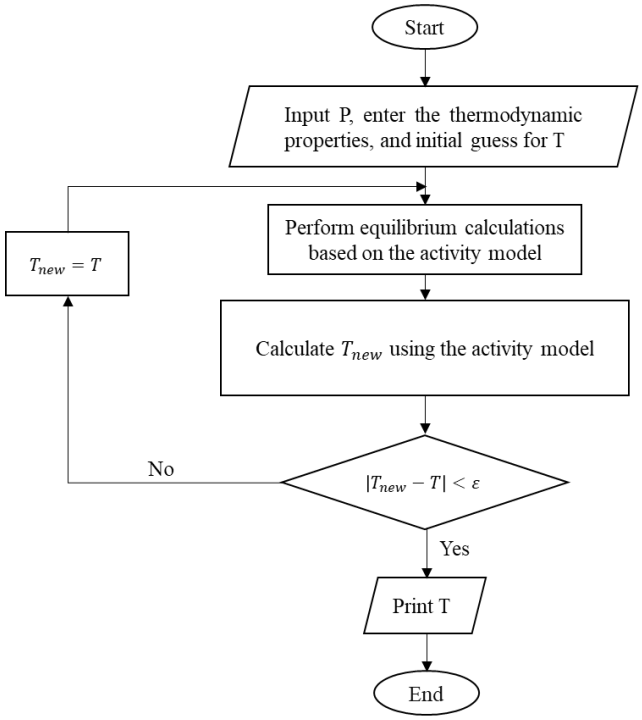


Figure 4-1: Flowchart of calculating hydrate dissociation pressure in a porous medium using the activity model.

4.2.2. Predictive models

The objective of this study is to accurately predict the dissociation temperature of CH_4 and C_3H_8 hydrates in porous media using input parameters such as pressure, pore diameter, and the critical properties of the gas (P_C and T_C). To achieve this, four different predictive models were developed: Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Multi-layer Perceptrons (MLP), all implemented in Python. The models were built using the Scikit-Learn library [79], a widely used and reliable Python library known for its flexibility and powerful features.

A Decision Tree is a non-parametric method that splits data into subsets based on input features, producing a tree-like structure where each decision node represents a rule, and each leaf node corresponds to a predicted outcome. It is considered a "white-box" model due to its visual interpretability. However, allowing the tree to grow too deep can lead to overfitting and reduce the model's generalization ability. A Random Forest is an ensemble technique composed of multiple Decision Trees. By averaging the predictions of several individual trees, it reduces the risk of overfitting associated with single trees and generally improves predictive performance. Both Decision Trees and Random Forests are non-distance-based models, relying instead on hierarchical feature splits. In contrast, a Support Vector Machine (SVM) finds the optimal hyperplane (or set of hyperplanes) in a high-dimensional space to separate data points or predict continuous outcomes. It is considered a "black-box" model because its internal decision-making process is less transparent. Similarly, a Multi-Layer Perceptron (MLP) is a type of neural network that uses multiple layers of interconnected neurons to capture complex, non-linear relationships within the data. Like SVM, MLPs are also considered "black-box" models and are distance-based. Further details on these methods can be found in the literature [79,80].

4.2.2.1. Data splitting

Proper data splitting is crucial for building reliable ML models, as it directly affects both accuracy and generalization. A well-structured split ensures enough training data to avoid underfitting, while keeping the test set separate helps evaluate the model's true performance and prevents overfitting. Consistent with approaches used in previous studies [157,158,196,197], in this study, we used a group-based splitting method instead of the common sample-based approach to improve the model's robustness, especially given the nature of hydrate dissociation data, which varies across gas types, pore sizes, and experimental conditions.

Unlike sample-based splitting, which randomly divides the dataset and may place related samples in both training and testing sets, group-based splitting keeps all samples from the same group in one set. This method helps prevent data leakage and provides a more accurate assessment of model performance. For instance, dissociation temperature is influenced by variables such as temperature, pore diameter, and gas type (e.g., CH_4, ethane, or carbon dioxide). If similar data

points appear in both training and testing sets, the model may produce overly optimistic results by identifying patterns it has already seen.

Group-based splitting is especially valuable in this study, where the aim is to capture general patterns across different gases and conditions, not just individual predictions. It pushes the model to generalize knowledge from one group to entirely new ones. As such, we used dissociation data for CH_4, ethane, and carbon dioxide across various pore sizes for training, while reserving C_3H_8 hydrate data for testing. Table 4-4 summarizes the dataset used to build the models, including gas types, ranges of pressure and temperature, pore sizes, and the number of samples.

Table 4-4: Summary of the data used for the training and testing sets, including gas types, pressure and temperature ranges, pore diameters, and sample counts.

Gas	Pore diameter (nm)	Number of data points	*T* (K)	P (MPa)	Reference
CH_4	6	11	269.65–280.95	2.84–10.50	[32,33]
	10	12	271.80–283.10	3.69–12.50	[34–36]
	15	14	273.30–285.60	3.30–13.31	[33,35,36]
	30	20	270.90–287.50	2.38–14.07	[32–35]
	50	4	278.50–281.60	4.80–6.70	[34]
	100	10	272.55–284.25	2.47–8.32	[32,37]
Ethane	6	5	271.97-278.07	0.975-2.529	[38]
	15	5	274.73-282.54	0.976-2.814	[38]
	30	5	277.2-284.4	0.903-2.505	[38]
	100	5	278.69-285.24	1.024-2.556	[38]
C_3H_8	6	4	263.53-264.86	0.233-0.309	[38]
	10	6	269.3-271.4	0.285-0.420	[37]
	15	5	268.48-272.58	0.187-0.401	[38]
	30	8	271.94–276.8	0.176–0.572	[37,38]
	50	3	276.1-277.5	0.401-0.609	[37]
	100	12	272.83–277.7	0.192–0.555	[37,38]
Carbon dioxide	6	10	269.88–277.65	1.40–3.89	[32,33]
	10	5	270.2–277.5	1.45–2.31	[35,37]
	15	11	272.30–279.70	1.45–3.92	[33,35]
	30	17	271.80–281.35	1.13–3.92	[32,33,35,37]
	50	4	275.2–282.7	1.65–4.60	[37]
	100	10	271.25–281.65	1.06–3.93	[32,37]

4.2.2.2. SHapley Additive exPlanations (SHAP) method

SHapley Additive exPlanations (SHAP) is a method used to understand how important each input feature is and how it affects the model's output. Unlike traditional techniques, SHAP does not rely on how the model is built but instead looks at the relationship between the inputs and the predictions. This makes it especially helpful for complex models like ensemble methods, where the exact prediction formula is not easy to see. Based on ideas from game theory, SHAP gives a clear and consistent way to explain the predictions of many different ML models [159,160].

4.3. Result and discussion

In this study, we estimated the dissociation temperatures of CH_4 and C_3H_8 hydrates in different porous media using two distinct approaches. To evaluate the accuracy of our model predictions, we calculated the Average Absolute Deviation (AAD) and Average Absolute Deviation Percentage (AAD%) for estimated temperature using the following equations, respectively:

$$AAD = \frac{1}{N_t}\sum_{i=1}^{N_t}\left|T_i^{cal} - T_i^{exp}\right| \tag{4-13}$$

$$AAD\% = \frac{100}{N_t}\sum_{i=1}^{N_t}\left|\frac{T_i^{cal} - T_i^{exp}}{T_i^{exp}}\right| \tag{4-14}$$

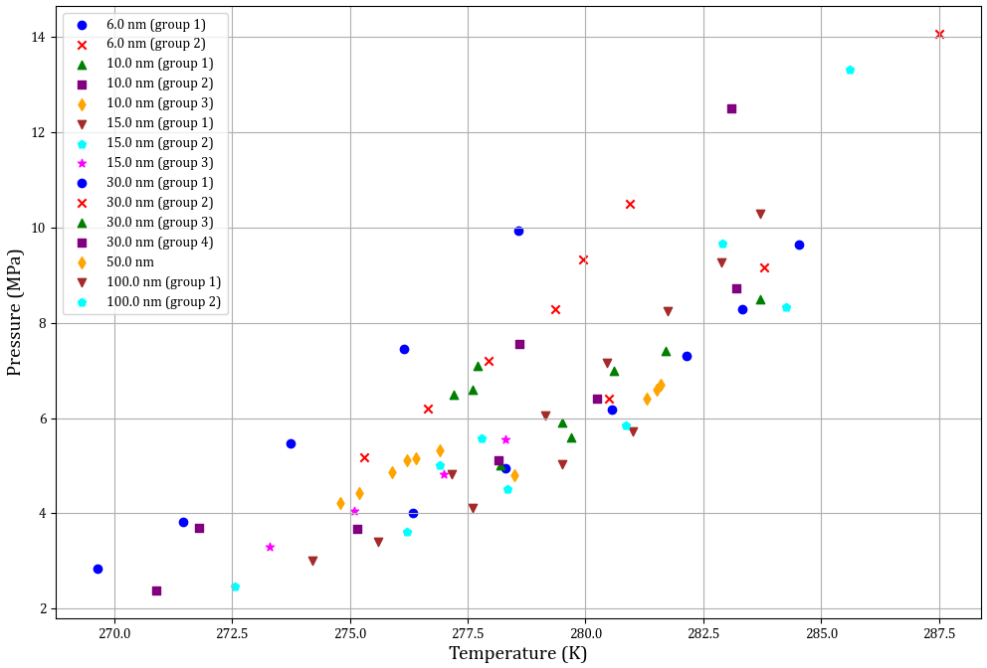
where N_t represents the total number of data points. This formula measures the average discrepancy between the calculated Temperatures (T_i^{cal}) and the experimental Temperatures (T_i^{exp}) over all data points, providing an indicator of how accurately the model predicts the experimental values.

Figure 4-2 provides a comprehensive overview of experimental CH_4 and C_3H_8 hydrate dissociation temperatures in various porous media, ranging from silica glass to silica gel, with pore diameters from 6 nm to 1000 nm. Each symbol corresponds to a specific pore size and/or experimental protocol, illustrating how systematic variations in pore diameter influence hydrate stability conditions. A key observation is that the dissociation boundary, as shown by the pressure–temperature (P–T) data points, shifts toward lower equilibrium temperatures in smaller pores

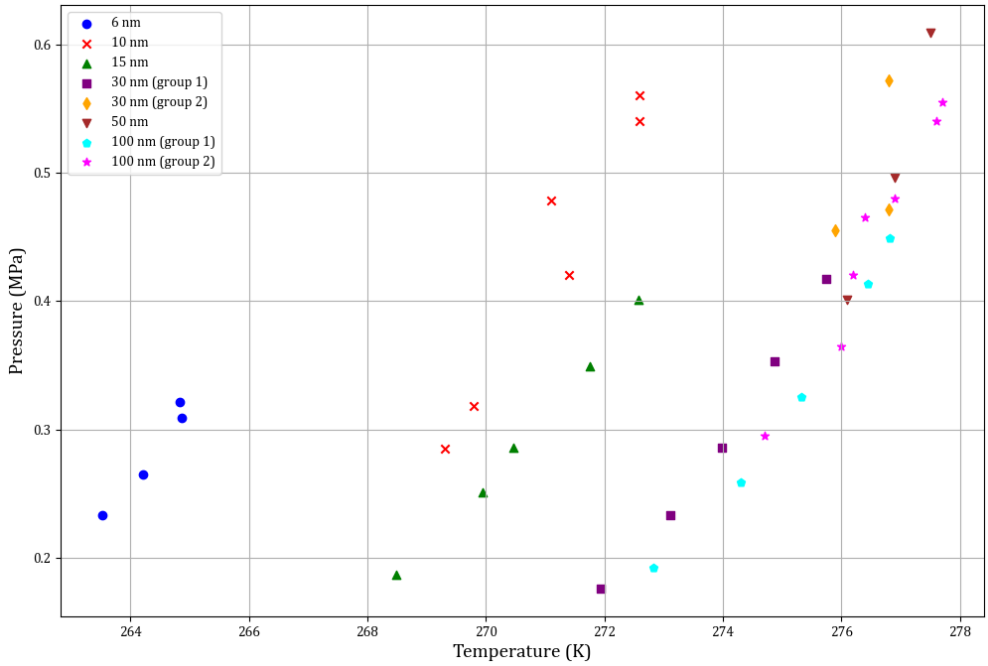
compared to larger ones at the same pressure. This trend is consistent with the Gibbs–Thomson effect, where capillary and surface forces stabilize the hydrate phase at lower temperatures. In smaller pores, the increased contact area between the hydrate and the hydrophilic silica surface amplifies these surface-driven effects, resulting in a progressive depression of the equilibrium boundary as pore size decreases [193].

In addition to the pore-size-dependent trend, the data exhibit noticeable scatter, which can be attributed to several sources. Differences in experimental procedures, including stepwise and continuous heating methods [198], can affect the observed dissociation point, with measurement errors also contributing to the variability. Furthermore, material synthesis plays an important role in data variability. For example, the scatter observed for 30 nm pores likely results from differences in material structure: silica gels [33] have irregular, amorphous pores, whereas Vycor glass [34] contains uniform, cylindrical pores. These structural differences influence pore connectivity and surface interactions, thereby affecting the dissociation behavior. Despite these sources of variability, all available experimental data points are retained to provide a transparent and comprehensive view of how experimental methods and porous media properties affect hydrate equilibrium. This inclusive approach contrasts with many modeling studies that rely on filtered datasets, and instead allows for a better appreciation of both the range of experimental uncertainty and the influence of pore-scale effects under diverse laboratory conditions.

Finally, it is worth noting that in larger pore systems, the dissociation temperatures begin to converge toward those of bulk hydrate systems. This transition suggests that capillary and interfacial effects diminish as pore size increases, leading to hydrate equilibrium behavior that more closely resembles that of the bulk phase [37]. Overall, as can be seen from Figure 4-2, the data highlights the complex and diverse nature of hydrate stability in confined environments, highlighting the importance of pore diameter, experimental technique, and material properties in determining hydrate dissociation conditions.



(a)



(b)

Figure 4-2: Overview of the experimental data used in this study showing equilibrium dissociation conditions of (a) CH_4 [32–37] and (b) C_3H_8 [37,38] hydrates across a range of porous media with different pore sizes.

4.3.1. Activity model

Figure 4-3 presents a comparison between the experimental data and the predicted CH_4 hydrate dissociation temperatures using the activity model in porous media with pore sizes ranging from 6 nm to 100 nm. Additionally, Table 4-5 shows the calculated AAD values for predicting the dissociation temperatures of both CH_4 and C_3H_8 hydrates in porous media using the activity model.

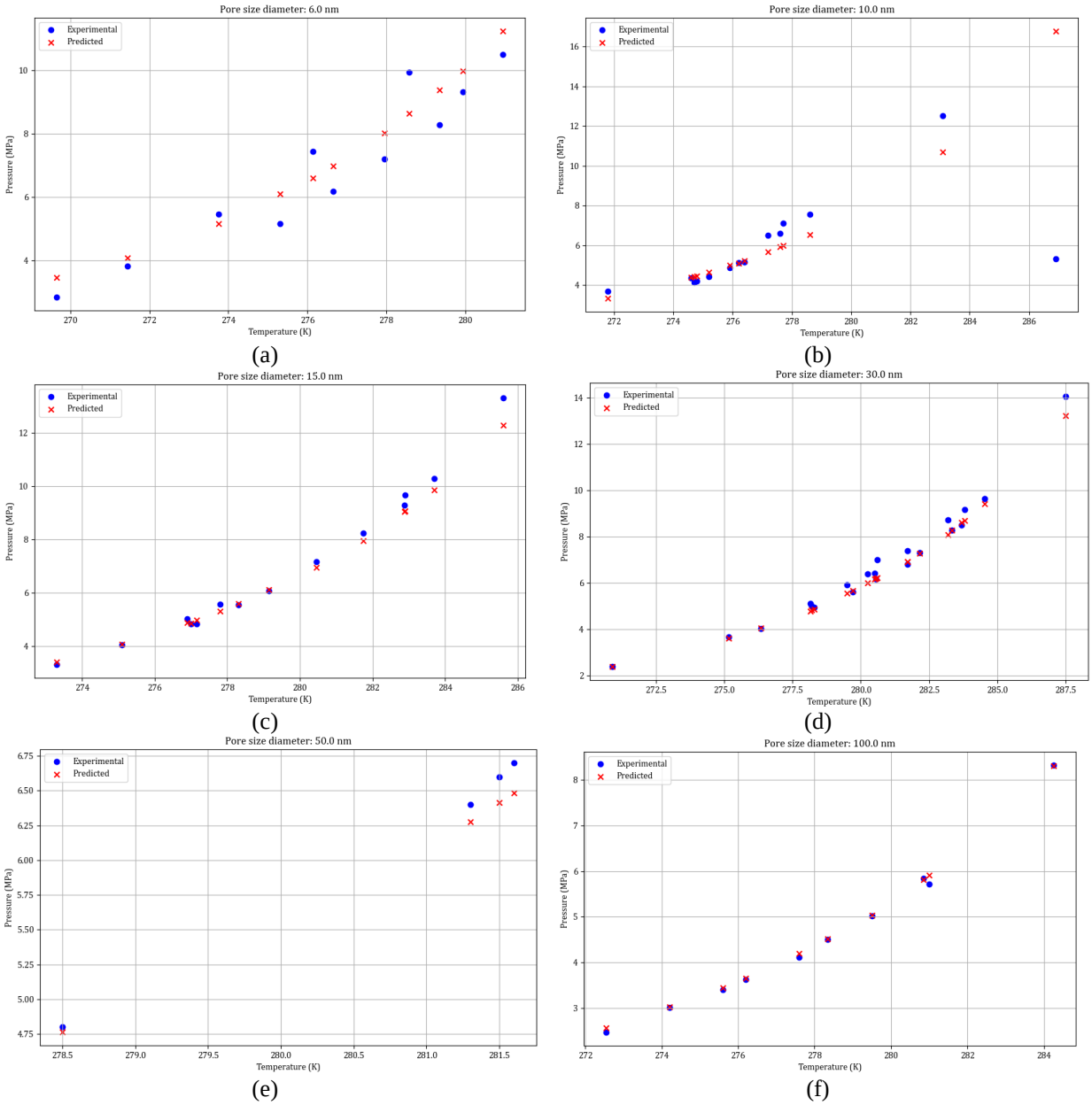


Figure 4-3: Experimental data and predicted CH_4 hydrate dissociation temperatures in porous media with different pore diameters: (a) 6 nm, (b) 10 nm, (c) 15 nm, (d) 30 nm, (e) 50 nm, and (f) 100 nm, using the activity model.

A surface tension value of 34 mJ/m^2 was used for the CH_4 hydrate–water interface, and 25 mJ/m^2 for the C_3H_8 hydrate–water interface. As can be seen from Figure 4-3, in smaller pores (6 nm, 10 nm, 15 nm), the experimental data exhibit significant shifts away from bulk hydrate dissociation conditions [181], caused primarily by capillary pressure and surface forces that become increasingly important as pore diameter decreases [37]. These shifts, which translate to hydrates remaining stable at higher pressures or lower temperatures compared to bulk conditions, are effectively captured by the model. Although the fit may show slightly larger deviations at the smallest pore sizes, likely due to real-world complexities such as surface heterogeneities, varying pore throat diameters, and local chemical effects [37,199], the model still captures the overall trend of a noticeable shift in the equilibrium curve. For, intermediate pores (30–50 nm), the model remains consistent with the experimental data. Its consideration of interfacial tension and water activity helps predict realistic dissociation points in this intermediate pore-size range. For, larger pores (approaching 100 nm), as pore size increases, the influence of curvature effects diminishes. The model appropriately reflects these weaker capillary forces, producing dissociation behavior that remains in line with observed results. Overall, the agreement across pore sizes from 6 nm to 100 nm suggests that the model effectively captures water–gas interfacial phenomena and core thermodynamic principles. This makes it a reliable tool for predicting hydrate dissociation in different porous materials.

Table 4-5: The calculated AAD for predicting the dissociation temperatures of CH_4 and C_3H_8 hydrates in porous media using the activity model.

Gas	Pore diameter (nm)	Number of data points	T-range /(K)	P-range / (MPa)	Porous medium	AAD	AAD%	Reference(s)
CH_4	**6**	5	269.65-278.58	2.84-9.94	Silica gel	1.056	0.383%	[32]
		6	275.30-280.95	5.17-10.5	Silica gel			[33]
	10	3	277.2-277.7	6.5-7.1	Silica glass	0.806	0.291%	[34]
		3	271.8-283.1	3.689-12.500	Silica glass			[35]
		6	274.8-276.9	4.14-5.32	Silica gel			[36]
	15	6	277.15-283.70	4.825-10.285	Silica gel	0.295	0.105%	[33]
		4	277.8-285.6	5.564-13.307	Silica glass Silica gel			[35] [36]

Gas	Pore diameter (nm)	Number of data points	T-range /(K)	P-range / (MPa)	Porous medium	AAD	AAD%	Reference(s)
			273.3-278.3	3.30-5.55				
	30	6 3 6 5	276.33-284.53 280.5-287.5 278.2-283.7 270.90-283.20	4.012-9.638 6.405-14.065 5.0-8.5 2.38-8.73	Silica gel Silica glass Silica glass Silica gel	0.360	0.128%	[33] [35] [34] [32]
	50	4	278.5-281.6	4.8-6.7	Silica glass	0.202	0.072%	[34]
	100	5 5	274.2-281.0 272.55-284.25	3.01-5.72 2.47-8.32	Silica glass Silica gel	0.118	0.043%	[37] [32]
	Total	71				0.473	0.170%	
C_3H_8	**6**	4	263.53-264.86	0.233-0.309	Silica gel	5.530	2.091%	[38]
	10	6	269.3-271.4	0.285-0.420	Silica glass	2.793	1.030%	[37]
	15	5	268.48-272.58	0.187-0.401	Silica gel	2.662	0.984%	[38]
	30	5 3	271.94-275.74 275.9-276.8	0.176-0.417 0.455-0.572	Silica gel Silica glass	0.449	0.163%	[37] [38]
	50	3	276.1-277.5	0.401-0.609	Silica glass	0.500	0.180%	[37]
	100	5 7	272.83-276.81 274.7-277.7	0.192-0.449 0.295-0.555	Silica gel Silica glass	0.554	0.200%	[37] [38]
	Total	38				1.682	0.624%	

Table 4-5 provides a quantitative evaluation of the activity model's performance through the calculated Average Absolute Deviation (AAD) and AAD percentage (AAD%) for CH_4 and C_3H_8 hydrate dissociation temperature predictions in porous media. For CH_4, the model demonstrates excellent agreement with experimental data across a wide range of pore sizes, achieving an overall AAD of 0.473 K and AAD% of 0.170%. As observed in the qualitative trends, the highest deviation occurs at the smallest pore size (6 nm), with an AAD of 1.056 K, likely due to

pronounced capillary and surface effects. However, the error steadily decreases with increasing pore diameter, reaching as low as 0.118 K at 100 nm, where the system approaches bulk-like behavior and curvature effects become negligible.

In contrast, the model's predictive performance for C_3H_8 hydrates, while still reasonable, is comparatively less accurate. The total AAD and AAD% for C_3H_8 were 1.682 K and 0.624%, respectively. The highest deviation was again observed at 6 nm (5.530 K), indicating greater modeling challenges under tight confinement for C_3H_8. Although accuracy improves with increasing pore size, the deviations remain higher than those for CH_4 across all diameters. This discrepancy may be attributed to propane's larger molecular size and stronger interaction with pore surfaces [185], as well as greater variability in the available experimental data.

Overall, the results reinforce the activity model's robustness in capturing pore-size-dependent hydrate behavior, particularly for CH_4. The consistently low AAD values in larger pores and the model's sensitivity to confinement effects confirm its suitability as a predictive tool for hydrate dissociation in porous materials.

When modeling the phase equilibrium conditions of gas hydrates in porous media, researchers come across multiple significant challenges. One major issue arises from the variability of experimental data. Different researchers often use dissimilar porous media, employ diverse experimental protocols, or rely on distinct measurement techniques. As a result, comparing or determining which dataset is the most reliable becomes difficult. Since the interfacial tension between hydrate and water is frequently obtained by regressing experimental phase equilibrium data, any unreliability in those experiments directly compromises the accuracy of the interfacial tension values, and consequently, the predictive power of hydrate equilibrium models [184].

Another important concern involves how porous media are represented in models. Some models simplify porous media by assuming that all pores are of a single size and therefore neglect the natural variation in pore size distributions [39]. In contrast, other approaches explicitly account for a range of pore sizes [191]. Both methods can produce reasonable results in certain scenarios. Because real porous media possess a wide variety of structures, synthesis methods, surface characteristics, porosities, and mechanical strengths. Experimental observations confirm that sediment characteristics such as composition, location, and particle size strongly influence hydrate phase equilibrium and stability. For instance, while the hydrate phase equilibrium temperature

remains fairly constant in coarse-grained sediments, it tends to decrease in fine-grained media (e.g., porous silica gel, glass sand, or marine sediment) relative to bulk water. Additionally, evidence suggests that CH_4 hydrate can form more readily in sands than in clay, highlighting how different sediment types exert distinct effects on hydrate behavior [185,200].

As discussed in the modeling section, various correlations [193] or constant values have been proposed for surface tension [35,37,201]. However, no single correlation consistently outperforms the others across all experimental datasets. Instead, each tends to work well within certain conditions but may fail when applied to different data. The same limitation applies to the equations of state used to calculate gas fugacity: each equation can yield accurate predictions for specific datasets but may show significant errors elsewhere [185]. Similar observations hold for activity-based [195], enthalpy-based [181], and fugacity-based models [202], each of which can excel when applied to the data for which it was developed but may not generalize effectively to different experimental conditions. These trends emphasize the general challenge that no single model or correlation performs equally well across all systems, highlighting the importance of validating predictive tools against multiple experimental datasets.

The current model assumes a zero contact angle ($\theta = 0$), which implies complete wetting of the hydrate phase by water. Although this simplification reduces computational complexity, it overlooks the influence of material-specific wettability. In reality, porous media vary in surface chemistry; for example, silica gel is hydrophilic, while organically modified silica is hydrophobic, both of which affect the actual contact angle. A nonzero contact angle changes the capillary pressure term $\frac{\gamma_{hl} cos\theta}{r}$ in Equation 4-5. On hydrophobic surfaces ($\theta > 0$), the value of $cos\theta$ decreases, which reduces the capillary pressure. This may result in underestimating dissociation conditions in hydrophobic media or overestimating them in hydrophilic systems where θ is incorrectly assumed to be zero. Incorporating measured contact angles or modeling θ as a function of surface properties can improve the representation of heterogeneous wettability in natural sediments.

In summary, modeling hydrate phase equilibrium in porous media demands careful selection of reliable experimental data, accurate representation of pore structures, and an awareness of how varying sediment characteristics affect equilibrium behavior. Researchers should also exercise caution when applying surface tension values/correlations, as their performance can vary

depending on system conditions and experimental variability. Addressing these challenges is essential for developing robust and universally applicable models of gas hydrate systems in porous media.

4.3.2. Predictive models

In this study, different machine learning (ML) algorithms were employed to predict the equilibrium dissociation temperature of gases based on their properties and operational parameters. The input features included the gas's critical temperature (T_C) and critical pressure (P_C), the pore diameter (representing the porous medium), and the system pressure (as an operational variable).

To ensure rigorous model evaluation and generalization across different gases, the dataset was partitioned by gas type rather than using random sampling. Specifically, all data related to C_3H_8 were excluded from the training set and reserved exclusively for testing. This gas-blind split strategy was adopted to assess the models' ability to generalize to previously unseen gases.

In addition to AAD and AAD%, residual (Res_i), coefficient of determination (R^2), root mean squared error (RMSE), and average relative deviation percentage (ARD%) were considered to evaluate the predictive performance of the ML models. The formulas for these metrics are provided below:

$$Res_i = t_i - y_i \tag{4-15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Res_i)^2}{\sum_{i=1}^{n}(\bar{t} - t_i)^2} \tag{4-16}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Res_i)^2} \tag{4-17}$$

$$ARD\% = \frac{100}{n} \times \sum_{i=1}^{n}\frac{Res_i}{t_i} \tag{4-18}$$

where y and t represent the model's output and target value, respectively.

Four ML algorithms were evaluated in this study: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). Key hyperparameters for each model were optimized using a randomized search strategy to identify the best-performing

configuration. Table 4-6 presents the search ranges, optimal hyperparameters, best cross-validation root mean squared error (RMSE), and corresponding training times.

Table 4-6: Optimal hyperparameters and performance metrics from randomized search.

Hyperparameter	Search Range	DT	RF	SVM	MLP
max_depth	3–50	39.0	11.0	–	–
max_features	0.01–1	0.265	0.244	–	–
n_estimators	4–300	–	17.0	–	–
C	0.1–1000	–	–	206.90	–
gamma	0.1–1000	–	–	0.104	–
hidden_layer_sizes	3–30	–	–	–	29.0
alpha	1e-5–1.0	–	–	–	0.0455
learning_rate_init	1e-5–0.1	–	–	–	0.0704
Best CV RMSE	–	0.3755	0.3087	0.1742	0.1762
Elapsed Time (s)	–	12.31	100.52	3.88	45.31

Among these models, the SVM and MLP, which are both distance-based learning algorithms, achieved the lowest cross-validation RMSE values, indicating strong predictive capabilities during training.

To ensure robustness and minimize the influence of randomness, each model was trained 100 times using different random seeds. The performance of the models was evaluated using multiple statistical metrics, including AAD, AAD%, ARD%, RMSE, and R^2, as summarized in Table 4-7. The results clearly show that the SVM model delivered the most consistent and accurate predictions on the test set, with an AAD of 1.4276 K, RMSE of 2.0932 K, and a high R^2 value of 0.7171. This indicates its strong ability to generalize to unseen data and accurately capture dissociation temperature trends across different gases.

Table 4-7: Metrics employed to measure the accuracy of the models developed in this work.

Model	Dataset	AAD	AAD (%)	ARD%	RMSE	R^2	N
DT	Training	0.000	0.0000	0.0000	0.0000	1.0000	148
DT	Testing	2.7471	1.0121	0.8275	3.2784	0.3060	38
RF	Training	0.3001	0.1081	0.0032	0.4330	0.9860	148
RF	Testing	2.2039	0.8176	0.5298	3.1415	0.3628	38
SVM	Training	0.3911	0.1409	0.0078	0.4930	0.9819	148
SVM	Testing	1.4276	0.5294	0.2683	2.0932	0.7171	38
MLP	Training	0.3130	0.1130	0.0036	0.4152	0.9872	148
MLP	Testing	12.2217	4.4808	4.4808	12.3091	-8.7832	38

In contrast, although the MLP model performed well during training, achieving the lowest training RMSE of 0.4152 K and the highest R^2 value of 0.9872, its test performance dropped significantly.

It exhibited the highest AAD of 12.2217 K, the highest ARD%, and a large negative R^2 value of −8.7832, indicating severe overfitting and poor generalization. The decision tree and random forest models also struggled with generalization, both yielding relatively high test RMSEs and low R^2 values. These outcomes suggest that while tree-based models may effectively capture patterns in the training data, they are less reliable when applied to new conditions, particularly in datasets with varying gas types and experimental settings.

Overall, the SVM model demonstrated the best balance between accuracy and generalizability, making it the most reliable tool for predicting hydrate dissociation temperatures in porous media based on the current dataset. The contrast between the MLP's strong training performance and poor test results highlights the importance of evaluating models not only on training metrics but also on their ability to generalize to unseen data, especially in systems characterized by high variability.

To interpret the SVM model's predictions and understand the contribution of each input feature to the predicted dissociation temperature, SHapley Additive exPlanations (SHAP) analysis was conducted. The beeswarm plot in Figure 4-4 provides a global overview of feature importance and its influence on the model output. Among the input features, critical temperature (T_C) exhibited the highest overall impact, followed by system pressure (P), critical pressure (P_C), and pore diameter. The plot reveals that higher values of T_C, pressure, and pore diameter are generally associated with higher predicted dissociation temperatures, indicating a positive correlation. In contrast, lower values of P_C contribute positively to the predicted dissociation temperature, suggesting a negative relationship between P_C and the model output.

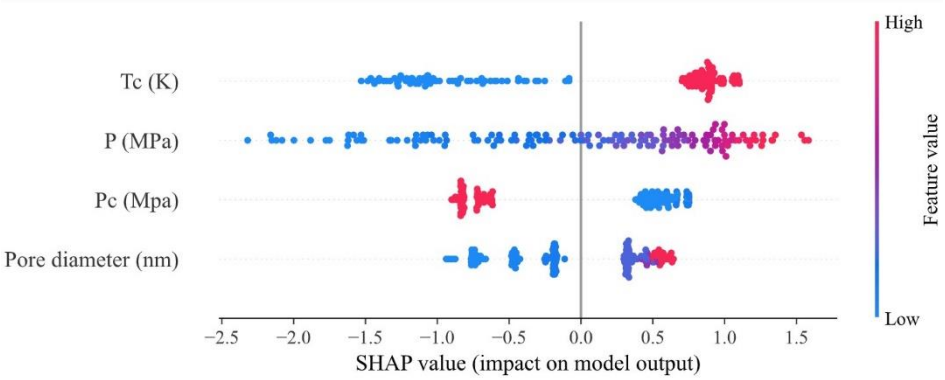


Figure 4-4: SHAP beeswarm plot showing feature impacts on SVM predictions.

As illustrated in the SHAP dependence plots in Figure 4-5, these relationships are clearly non-linear. The plots for T_C and P_C show distinct clustering of SHAP values, reflecting the discrete nature of these gas-specific properties in the dataset. This is expected, as T_C and P_C are fixed for each gas, and the model was trained using data for CH_4, ethane, and carbon dioxide, while data for C_3H_8 was used exclusively for testing. Although the use of group-based data splitting improved the model's ability to generalize to unseen gas types by preventing information leakage, the limited number of gases in the dataset still results in stepwise behavior in these features. This highlights that, while the model can generalize within the scope of gases it has seen, its ability to extrapolate to gases with significantly different critical properties could be enhanced by including a wider variety of gases in the training process.

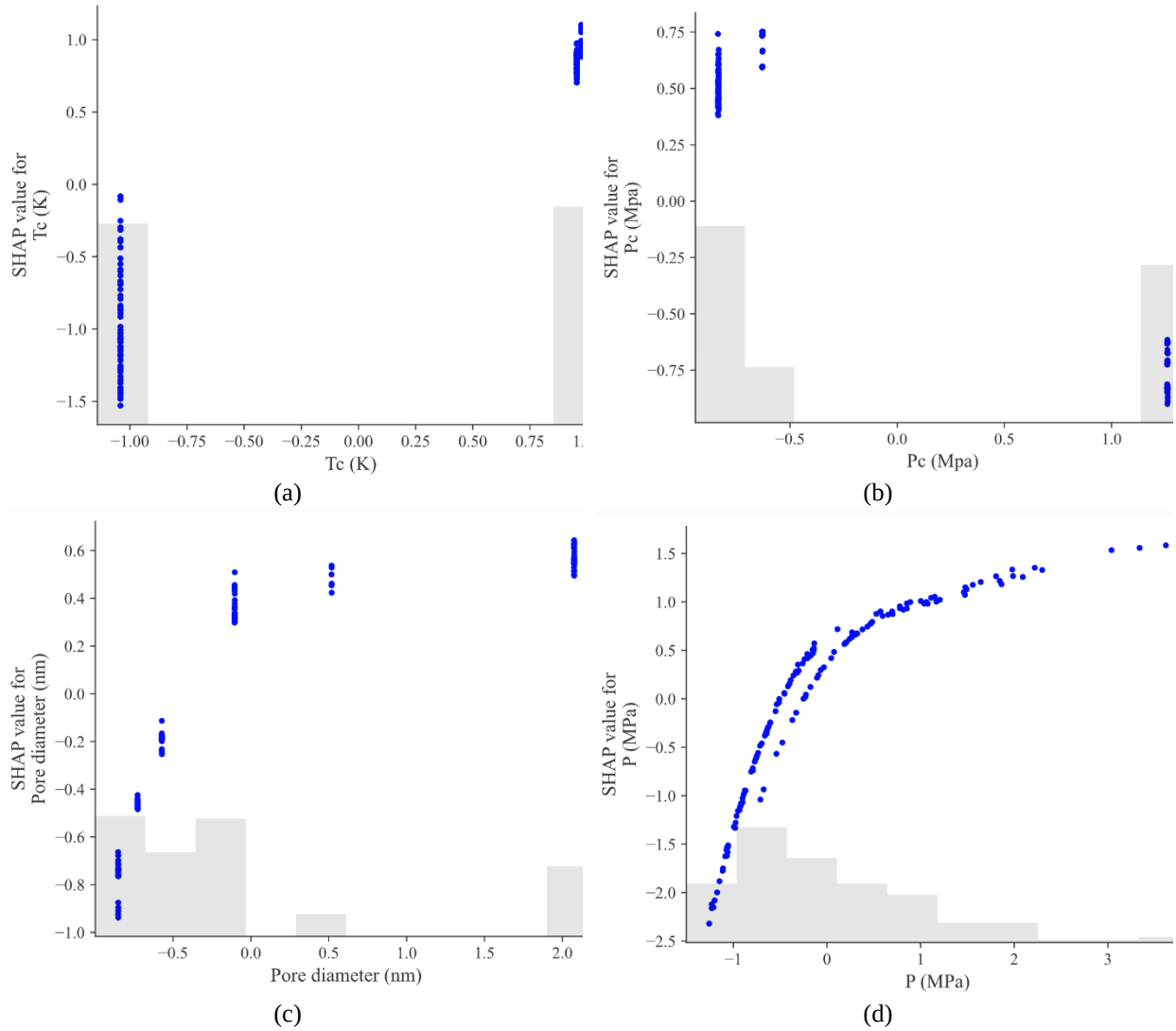


Figure 4-5: Scatter plots of input features vs. predicted equilibrium temperature.

In contrast, the plots for pressure and pore diameter show smoother and more continuous trends. The pressure feature exhibits a strong, logarithmic-like increase in SHAP values, indicating that higher system pressures have a substantial positive impact on the predicted dissociation temperature. Pore diameter shows a weaker but still consistent trend, with smaller pores contributing negatively and larger pores contributing positively to the prediction. This aligns with the physical understanding that confinement in smaller pores reduces hydrate stability, resulting in lower dissociation temperatures.

In conclusion, the SHAP analysis demonstrates that the SVM model captures key non-linear relationships and produces physically meaningful predictions for dissociation temperature. While the model benefits from group-based splitting for assessing generalization, the findings also suggest that expanding the diversity of gases in the training dataset would improve the model's robustness and its ability to generalize to a wider variety of hydrate systems.

The use of ML models can offer a promising approach to overcoming the limitations of traditional thermodynamic models, resulting in better performance across diverse data sources. Beyond the inputs considered in this study, ML techniques can simultaneously incorporate additional influential factors such as surface properties and the structural characteristics of porous materials. This allows for the development of more comprehensive and accurate models for predicting the phase equilibrium of gas hydrates in porous media.

4.4. Conclusions

The study employs both thermodynamic and machine learning techniques to model the dissociation behavior of gas hydrates across a range of pore sizes in porous materials. Using the activity model, which accounts for the effects of confinement on hydrate dissociation, we successfully predicted the dissociation conditions of CH_4 and C_3H_8 hydrates. The model performed with high accuracy across all pore sizes, particularly for methane, achieving a total AAD% of 0.17%, and demonstrated its ability to reflect confinement effects, especially at small pore diameters.

For the machine learning component, four algorithms were compared using a gas-blind data splitting strategy. The SVM model consistently outperformed others, achieving an AAD of 1.43

K and an R^2 of 0.72 on the test set consisting exclusively of C_3H_8 data. SHAP analysis provided physical insight into feature importance and confirmed that higher critical temperatures, pore diameters, and system pressures positively influence dissociation temperature, while critical pressure shows a negative correlation. These trends align with physical expectations and reinforce the model's interpretability.

However, limitations remain. The activity model assumes a zero contact angle and uses fixed interfacial tension values, which may not fully capture variations in surface chemistry or wettability. Similarly, the ML models, while generalizable, are constrained by the limited diversity of gas types in the training dataset, as evident from discrete SHAP patterns for gas-specific features. Expanding the dataset to include additional gases and porous materials will be essential for further improving the model's robustness and generalization.

In addition to the current study, future research on predicting the equilibrium dissociation conditions of gas hydrates under confined conditions can explore several promising directions. Expanding the dataset to include a wider range of conditions and materials, such as different types of porous media, can help improve model accuracy. Adding details about porous media properties, such as pore size distribution and surface roughness, to the experimental data and model inputs can also make predictions more reliable. Exploring advanced ML techniques, including deep learning, may further improve accuracy, especially in complex systems. These efforts build on the foundation of this study and can support better approaches for predicting the phase equilibrium of gas hydrates in porous media.

In summary, combining thermodynamic and ML-based methods provides a comprehensive and scalable framework for predicting gas hydrate behavior in confined systems. These models hold promise for applications in energy resource development, carbon capture, and environmental risk assessment.

Chapter 5: Investigation of hydrate particle dynamics and distribution in water-dominated flows

5.1. Introduction

Gas hydrates, also called clathrate hydrates, are solid crystals formed when water molecules create cage-like structures through hydrogen bonding, trapping gas or volatile liquid molecules inside. These cages stabilize the guest molecules and can form three main structures: structure I (sI), structure II (sII), or structure H (sH), depending on the size of the trapped molecules. The structure type affects the formation and natural occurrence of gas hydrates [2].

Natural gas hydrates hold promise as an emerging energy resource due to their wide distribution and large potential. Production systems often use separate gas and water lines to reduce hydrate risks, but gas can still enter water lines, causing hydrate formation. The behavior of hydrates in these water-rich flows is not well understood [203]. This build-up can significantly reduce the flow area, restricting fluid movement and increasing the risk of blockages [109] and hinder operational efficiency [55,161,204]. Furthermore, deepwater drilling operations are vulnerable to gas hydrate formation due to the inherently low temperatures and high pressures in these environments [205]. To sum up, the formation, growth, aggregation, and deposition of hydrates within flowlines greatly increase the risk of blockages, making it essential to analyze and better understand these processes [40]. In addition to providing clearer insight into hydrate blockages, experimental investigations have led to the development of numerous physical models describing solid hydrate blockage phenomena.

Wang et al. [206] conducted hydrate formation experiments in a high-pressure, fully visual flow loop using a pure water system to simulate multiphase flow. The 50 m-long loop allowed detailed observation of hydrate behavior under varying liquid loading and pump speeds. At low liquid loading, two regions were observed: stable pressure with high formation, followed by pressure drop and reduced growth. At higher loadings, three regions appeared: stable pressure with peak formation, rapid pressure rise with slower growth, and fluctuating pressure as formation declined. Liu et al. [207] investigated hydrate deposit evolution in water-dominated systems using high-pressure reactors and Focused Beam Reflectance Measurement (FBRM). Higher hydrate

concentrations formed layered deposits, while lower ones resulted in suspended slurries. When moisture content exceeded 87%, downward pore water flow drove deposit growth. Larger particles had greater permeability and migration rates, and uniform particle sizes helped maintain suspension. A Darcy's law-based model was developed and matched well with experimental data, confirming the porous nature of hydrate beds. Kong et al. [208] conducted hydrate formation experiments in a water-dominated bubbly flow using a horizontal high-pressure flow loop. Using FBRM, they tracked particle sizes and analyzed flow and blockage behavior. They identified four stages of flow barrier formation, with large aggregates (150–500 μm) driving blockage. A modified kinetic model was developed for cooling conditions, and a critical flow velocity (0.65 ± 0.07 m/s) was determined for large aggregate formation. Two blockage types were observed: sedimentary beds at low flow rates and stationary beds at higher ones. Yao et al. [209] studied the growth and stability of hydrate deposit beds at the gas–liquid interface in water-dominated systems using stirred reactors and high-speed cameras. They found that hydrate films formed at the interface and evolved into bedding layers as particles settled from the liquid phase. Peak bed growth rates exceeded 0.15 mm/min, influenced by subcooling and shear rate. At hydrate volume fractions of 1–3%, a bilayer structure could form. Based on their results, they developed models to describe hydrate bed growth and predict critical Reynolds numbers and breaking speeds. Sakurai et al. [210] conducted flowloop experiments to study hydrate formation rates and slurry viscosity in water-dominant systems. Using a 10 mm ID flowloop at 100% water cut, they captured high-resolution images of hydrate growth in turbulent bubbly flow with an in-line camera. They developed a mass-transfer-limited growth model that accounts for changes in viscosity and gas transfer, incorporating small eddy cell theory and differential effective medium theory. The gas-water interfacial area was directly measured from images. In a subsequent study, Sakurai et al. [203] investigated methane hydrate particle dynamics in water-dominated systems under transient flow using a flowloop with an in-line video camera. They examined hydrate behavior during a sudden stop and stepwise flow reduction. Agglomeration occurred rapidly, altering particle size distributions within seconds. Larger, irregular aggregates formed under lower shear, indicating higher slurry viscosity. These results supported a new conceptual model for hydrate blockage, providing insights into microscale behavior and re-association risks in hydrate production and CO_2 sequestration. In a recent study, Sakurai et al. [211] performed flowloop experiments to examine the rheology of methane hydrate slurries in a 100% water cut system, focusing on the effects of

hydrate volume fraction on pressure drop, viscosity, and yield stress. Using a video camera and applying a Bingham plastic model, they found that pressure drop remained stable below ~14 vol% hydrate but increased sharply beyond that, linked to the turbulent–laminar flow transition. The model accurately described behavior up to 19 vol%, though wall slip affected results at higher concentrations. Even at low hydrate levels (~3.7 vol%), cohesive aggregates filled the pipe, highlighting their distinct behavior from noncohesive suspensions. Wang et al. [212] developed a transient simulation model to predict hydrate formation rates in both oil- and water-dominated pipeline systems under deepwater multiphase flow conditions. The model incorporates different mechanisms for each system type: interfacial formation in oil-dominated flows and mass transfer-driven growth in water-dominated flows. Validated through pilot-scale flowloop experiments, it was also applied to a subsea tieback case. The model accounts for factors like dispersion inversion and pipeline inclination on hydrate accumulation.

Compared to experimental methods, numerical simulation offers clear advantages, such as saving resources and enabling detailed parameter sensitivity analysis. Computational Fluid Dynamics (CFD) is widely used as a modern tool for modeling changes in fluid-related physical fields. By applying discretization techniques, continuous fields are represented by a finite number of variables, allowing efficient simulation and analysis of complex fluid dynamics phenomena [40].

Rensing et al. [213] developed a 3D Discrete Element Method (DEM) model coupled with CFD to simulate hydrate particle aggregation and structural evolution, which are challenging to observe experimentally. The model captures particle motion and fluid-particle interactions, enabling analysis of how shear, viscosity, and interparticle forces influence hydrate structure and flow. It also estimates slurry viscosity and compares it to hard sphere theory, enhancing understanding of hydrate plugging in flow systems.

Song et al. [42] developed a population balance-based dynamic model to simulate hydrate particle agglomeration and breakage in pipelines. Using CFD (ANSYS FLUENT), they examined the effects of flow rate and hydrate volume fraction on agglomeration. The model included factors such as collision frequency and particle size distribution, offering insights to aid hydrate risk management in offshore pipeline systems.

Liu et al. [214] developed a Two-Fluid Model (TFM) for hydrate slurry flow that incorporates inter-particle adhesion, enhancing accuracy compared to traditional models. Coupled with a

population balance model, it simulated variations in particle size and number, showing that adhesion increases mean particle size and strongly influences flow behavior. The model also examined the effects of different inlet boundary conditions, offering insights into adhesive liquid-solid flow dynamics.

Considering the interaction forces and dynamic behavior of particles in two-phase flows, the use of a CFD–DEM coupling strategy allows for more accurate and efficient predictions [215]. Hellestø et al. [216] applied the DEM with a soft-sphere collision model to simulate cohesive particle agglomeration in shear flow. They developed custom post-processing tools to assess agglomerate size, number, and fractal characteristics. The results showed that higher shear rates promoted agglomeration, though particle size eventually stabilized. Over time, weaker agglomerates broke apart, while stronger and denser ones formed.

Blais et al. [217] developed an unresolved CFD–DEM model to simulate viscous solid–liquid suspensions, with a focus on laminar and transitional mixing. The model couples fluid and particle dynamics and incorporates a sub-grid viscosity model to improve rheological accuracy. It was applied to a stirred tank with a pitched blade turbine and validated against experimental data on particle distribution and suspension behavior. Duan et al. [40] applied a coupled CFD–DEM approach to simulate the hydrate blockage process in a water-dominated pipeline system. Their 3D model captured hydrate particle behavior, including aggregation, deposition, and collisions, while accounting for cohesion between particles and adhesion to pipe walls. They explored two blockage scenarios: wall deposition and aggregate collisions. Critical blockage states were identified based on changes in hydrate volume fraction and pressure drop, providing insights into how hydrate buildup leads to flow restriction.

Wang et al. [218] employed a CFD–DEM model to simulate the effects of hydrate adhesion properties such as agglomeration, cohesion, and deposition on flow safety during solid fluidization exploitation of marine gas hydrates. They found that adhesion forces promote stationary bed formation, raise the critical deposition velocity, and reduce the height of the moving bed. Particle accumulation at the pipe bottom decreased the effective flow area, emphasizing the importance of careful design in hydrate slurry transport systems. Zhang et al. [43] used a coupled CFD–DEM model with ANSYS Fluent and EDEM to simulate hydrate particle flow in a water-dominated horizontal pipeline. At low flow rates, pressure drop was mainly caused by particle agglomeration

and sedimentary bed formation, while at high flow rates, it was due to fluid-wall friction. Particle velocities were lower near the pipe wall and higher at the center, providing insights into hydrate transport and blockage behavior. Wang et al. [44] used a coupled CFD–DEM model with ANSYS Fluent and EDEM to simulate hydrate aggregation and deposition in gas-dominated pipelines. They found that adhesion forces, particularly tangential ones, play a central role in particle aggregation. Deposition was strongly influenced by factors such as subcooling, particle size, and flow rate, and the study emphasized the importance of coordination number in assessing aggregation behavior.

In this study, a two-way coupled CFD–DEM framework is employed to investigate the dynamic behavior of hydrate particles in water-dominated horizontal pipe flow. The analysis covers a range of inlet velocities and particle volume fractions to assess the impact of flow conditions on interphase momentum exchange, particle–particle interactions, and spatial distribution. Time-resolved evaluations of pressure drop, drag force, particle velocity, and radial migration provide detailed insight into flow regime transitions, mechanical resistance, and particle transport mechanisms. The findings contribute to a deeper understanding of hydrate slurry behavior in multiphase pipeline systems and support the development of more accurate predictive models for flow assurance applications.

A key contribution of this work is the detailed, time-resolved characterization of hydrate particle dynamics during multiphase transport, achieved through a coupled CFD–DEM approach based on an open-source software package. Utilizing OpenFOAM and LIGGGHTS within the CFDEM® coupling framework, the study integrates quantitative metrics with cross-sectional velocity visualizations to reveal the early evolution of clustering, suspension dynamics, wall interactions, and particle depletion across different concentrations and flow conditions. The results suggest that volume fraction and flow velocity play important roles in hydrate transport behavior and offer useful insights for modeling and managing hydrate risks in subsea pipeline systems.

5.2. Methodology

This study adopts a two-way coupled CFD–DEM (Computational Fluid Dynamics–Discrete Element Method) approach to simulate hydrate particle transport in water-dominated horizontal

pipe flow. The simulation is carried out using the open-source CFDEM® framework, which combines OpenFOAM for fluid dynamics with LIGGGHTS for discrete particle modeling. This integrated approach captures the complex interactions between the continuous water phase and discrete hydrate particles through interphase momentum exchange mechanisms. These include drag force, pressure gradient force, and viscous force, each of which plays a critical role in accurately simulating particle motion and fluid–solid coupling.

5.2.1. Fluid phase modeling

The fluid phase dynamics are governed by the fundamental principles of mass and momentum conservation, described by the Navier–Stokes equations. These equations account for the behavior of the liquid (water) as the continuous phase in the solid–liquid flow system. The continuity equation ensures the conservation of mass within the control volume, while the momentum equation describes the balance of forces acting on the fluid, including pressure gradients, viscous stresses, gravitational forces, and momentum exchange with the solid phase. Mathematically, the conservation of mass and momentum are expressed as follows, respectively [43,219,220]:

$$\frac{\partial}{\partial t}(\varepsilon_l \rho_l) + \nabla.(\varepsilon_l \rho_l u_l) = 0 \tag{5-1}$$

$$\frac{\partial}{\partial t}(\varepsilon_l \rho_l u_l) + \nabla.(\varepsilon_l \rho_l u_l u_l) = -\varepsilon_l \nabla \mathrm{p} + \nabla.(\varepsilon_l \tau_l) + \varepsilon_l \rho_l g + S \tag{5-2}$$

where ε_l denotes the liquid volume fraction, ρ_l the liquid density, $u_l$the velocity field, p the pressure, τ_l the viscous stress tensor, g the gravitational acceleration, and S the momentum exchange source term resulting from fluid–particle interactions. The interaction term S captures the cumulative influence of drag and other forces exerted by suspended particles on the surrounding fluid.

$$S = \frac{1}{\Delta V}\sum_{i=1}^{n}(F_D)_i \tag{5-3}$$

where ΔV is the control volume, n is the number of particles within that volume, and $(F_D)_i$ is the drag force acting on the ith particle. This term quantifies the cumulative impact of fluid–particle interactions on the momentum balance in the system.

5.2.2. Particle motion modeling

The motion of hydrate particles within the fluid is governed by Newton's second law, which accounts for both translational and rotational dynamics [43,221]. Each particle is treated as an individual discrete entity, and its trajectory is determined by the balance of forces acting upon it. The translational motion of each particle is expressed as [43,222]:

$$m_i \frac{du_p}{dt} = m_i g + f_{p-l,\,i} + \sum_{j=1}^{k_i} f_{contact,\,ij} \tag{5-4}$$

where m_i is the mass of particle i, u_p is its velocity, g is gravitational acceleration, $f_{p-l,\,i}$ is the fluid force acting on the particle, and $f_{contact,\,ij}$ represents the contact force between particles i and j. The summation accounts for all k_i particles in contact with particle i. The rotational motion of the particle is given by:

$$I_i \frac{d\omega_i}{dt} = \sum_{j=1}^{k_i} T_{ij} \tag{5-5}$$

where I_i is the moment of inertia, ω_i is the angular velocity, and T_{ij} denotes the torque arising from interactions with neighboring particles. These equations comprehensively describe how hydrate particles respond to fluid dynamics and inter-particle interactions within the CFD-DEM simulation environment.

5.2.3. Fluid–particle interaction forces

As hydrate particles move within the fluid, they are subjected to several interphase forces that arise due to relative motion and differences in physical properties between the solid and liquid phases. The accurate representation of these forces is essential for capturing realistic particle behavior and transport phenomena.

The model employed in this study incorporates three key force contributions: pressure gradient force [223], viscous force [224], and drag force as described by Gidaspow's model [225]. These forces are computed dynamically within the simulation framework and directly influence the particle momentum equations, contributing to the source term in the fluid momentum equation. Together, these interaction forces determine the response of particles to the surrounding fluid field, affecting acceleration, rotation, and clustering behavior. Their integration into the CFD–DEM model aims to capture essential aspects of multiphase dynamics in hydrate transport systems, recognizing inherent modeling simplifications.

5.2.4. Simulation setup and assumptions

To investigate hydrate particle transport, a representative flow configuration is established. The computational domain consists of a horizontal pipe with a circular cross-section having a diameter of 0.024 m and an overall length of 0.5 m. Hydrate particles are modeled as spherical and monodisperse with a uniform diameter of 0.0005 m. A schematic of the computational domain and hydrate particle distribution is shown in Figure 5-1.

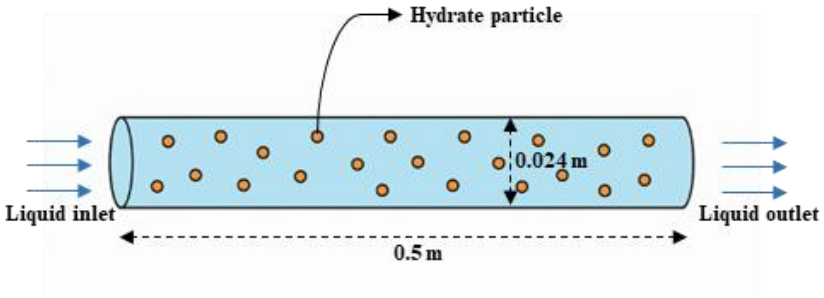


Figure 5-1: Schematic of the computational domain with hydrate particle distribution.

A transient simulation is carried out using a two-level time stepping scheme: a CFD time step of 0.00004 s governs the fluid solver, while a finer DEM time step of 0.0000004 s allows for detailed resolution of particle collisions and interactions. One hundred DEM sub-steps are performed for each CFD time step to maintain strong coupling fidelity.

The Eulerian–Lagrangian multiphase flow model is used to simulate the interactions between the continuous and discrete phases. In this framework, water is treated as the continuous (Eulerian) phase, while hydrate particles are individually tracked as discrete entities using a Lagrangian approach. This modeling strategy enables accurate resolution of interphase momentum exchange and particle-scale dynamics. The fluid flow is resolved using the PISO algorithm, with turbulence effects captured using the RANS k–ε model. Gidaspow's drag force model is employed in this study to capture the fluid-particle interactions, and additional forces, such as the pressure gradient and viscous force, are also included. Particle–particle and particle–wall interactions are modeled using the Hertz–Mindlin contact model with SJKR cohesion [226], which accounts for surface energy effects during contact.

Boundary conditions are specified for both fluid and particle phases. For the fluid (CFD), pipe walls are treated as no-slip boundary condition with zero-gradient pressure; the inlet applies a fixed velocity and zero-gradient pressure, while the outlet is defined as internal pressure with zero-gradient velocity. Turbulence variables at the walls employ wall functions, and at the inlet and outlet, fixed or zero-gradient conditions are applied as appropriate. For the particle phase (DEM), hard wall boundaries are enforced and no additional particles are injected during the simulation runtime. The water is assumed to be incompressible and isothermal, and the hydrate particles are treated as rigid, non-deforming spheres. Thermal effects, phase changes, and chemical reactions are not considered.

To represent hydrate particle formation, a virtual particle factory is introduced in the simulation. This factory, modeled as a cylinder matching the pipe dimensions, serves as a region where particles are randomly generated to approximate hydrate formation. The particles enter the flow at the same velocity as the continuous phase, ensuring downstream transport by water flow. Key assumptions of this model include: 1) hydrate formation is approximated by random particle generation within the virtual factory; 2) all hydrate particles are uniform in size and perfectly spherical; and 3) particle integrity is maintained throughout the simulation with no breakage or deformation. While this approach simplifies the complex formation process, it provides a practical means to study particle transport dynamics.

The physical and mechanical properties used in the simulation are listed in Tables 5-1 and 5-2, which define both fluid characteristics and particle behavior under flow conditions.

Table 5-1: Fluid properties.

Property	Value
Fluid Density (kg/m^3)	1000
Fluid kinematic viscosity (m^2/s)	1×10^{-6}

Table 5-2: Particle properties.

Property	Value
Poisson's ratio	0.45
Friction Coefficient	0.05
Restitution Coefficient	0.9
Density (kg/m^3)	960

5.2.5. Coupling strategy

The CFD–DEM coupling process is executed through a tightly integrated sequence of operations that allow fluid and particle phases to interact dynamically:

1) Initialization: The CFD solver initializes the fluid field, and the DEM solver assigns initial particle positions and velocities.
2) Porosity and Cell Mapping: Particles are mapped to fluid cells, and porosity (solid volume fraction) is computed.
3) Force Calculation: Fluid–particle interaction forces (drag, pressure gradient, etc.) are calculated and applied to the corresponding particles.
4) DEM Time Integration: Particle positions and velocities are updated using both contact and fluid interaction forces across multiple DEM sub-steps.
5) Data Exchange: Updated particle information is passed back to the CFD solver, which recalculates porosity and momentum source terms for the next time step.
6) Time Marching: This process is repeated until the simulation reaches the desired final time.

5.3. Results and discussion

This study examines the pressure drop behavior of hydrate particle–water mixtures flowing through a horizontal pipe. The dispersed phase consists of hydrate particles—solid, rigid, and spherical—suspended in a continuous water phase. Simulations were conducted using a CFD–

DEM framework at three inlet superficial velocities (0.2, 0.5, and 0.8 m/s) and five hydrate volume fractions (2%, 5%, 8%, 15%, and 20%) for each velocity.

5.3.1. Pressure drop response to hydrate loading and flow rate in horizontal pipe flow

To quantify flow resistance, average pressure values were extracted at two axial locations (z = 0.10 m and z = 0.49 m) along the pipe, from which the pressure drop (ΔP) and corresponding pressure gradient (ΔP/L) were calculated. These values were averaged over the early part of the simulation (the first 50 saved intervals, with data written every 0.004 seconds), equating to a total of 0.2 seconds of simulated flow time. This early-time averaging window captures the initial, quasi-steady hydrodynamic response before the onset of particle migration, clustering, or outlet loss. By focusing on this initial period, the results provide a clean snapshot of flow resistance induced by hydrate loading and avoid artifacts introduced by longer-term particle behavior. The results are summarized in Table 5-3.

Table 5-3: Averaged pressure drop and pressure gradient (ΔP/L) for hydrate–water mixtures at varying flow rates and volume fractions.

No.	Volume Fraction (%)	Number of Particles	Flow Rate [m/s]	Avg Pressure @ z = 0.10 m [Pa]	Avg Pressure @ z = 0.49 m [Pa]	Pressure Drop (P1 − P2) [Pa]	ΔP/L [Pa/m]
1	2	8639	0.2	0.30453	0.01453	0.29000	0.58000
2	5	21600	0.2	1.45673	0.05723	1.39951	2.79900
3	8	34600	0.2	0.51594	0.04016	0.47578	0.95156
4	2	8639	0.5	0.21712	0.01004	0.20708	0.41416
5	5	21600	0.5	0.74736	0.45729	0.29008	0.58014
6	8	34600	0.5	1.50567	0.05058	1.45508	2.91018
7	2	8639	0.8	0.50190	0.02244	0.47946	0.95892
8	5	21600	0.8	1.89500	0.05379	1.84121	3.68242
9	8	34600	0.8	3.42634	0.14675	3.27959	6.55918
10	15	64800	0.2	0.80559	0.03271	0.77288	1.54576
11	20	86400	0.2	3.33861	0.11575	3.22286	6.44572
12	15	64800	0.5	4.32633	0.25045	4.07588	8.15176
13	20	86400	0.5	8.88951	0.31113	8.57838	17.15676
14	15	64800	0.8	11.91231	2.14986	9.76246	19.52490
15	20	86400	0.8	12.16124	1.71202	10.44922	20.89844

The data reveal that, at fixed flow rates, the pressure drop increases significantly with particle concentration, highlighting the growing influence of interphase interactions and wall friction. At a superficial velocity of 0.2 m/s, the pressure gradient (ΔP/L) rises from 0.58000 Pa/m at 2%

volume fraction to 6.44572 Pa/m at 20%. This largely monotonic trend, with only minor variations, suggests particle accumulation and layering effects within the pipe cross-section, which contributes to increased shear drag. At 0.5 m/s, ΔP/L ranges from 0.41416 Pa/m at 2% to 17.15675 Pa/m at 20%. The sharp increase beyond 8% indicates a likely transition into a denser, pseudo-slurry regime, where particle–particle interactions become the dominant source of resistance. Similarly, at the highest velocity of 0.8 m/s, ΔP/L increases from 0.95892 Pa/m at 2% to 20.89844 Pa/m at 20%. The shape and steepness of this curve suggest that while higher velocities promote particle suspension, the accompanying rise in inertia and turbulence significantly amplifies energy loss through interphase drag and flow resistance.

The influence of flow rate is also evident when the volume fraction is held constant. At 2% volume fraction, ΔP/L increases from 0.58000 Pa/m at 0.2 m/s to 0.95892 Pa/m at 0.8 m/s, though a slight dip to 0.41416 Pa/m at 0.5 m/s may indicate a temporary suspension effect that reduces near-wall interactions. For 5% volume fraction, the pressure gradient follows a consistent upward trend from 2.79900 to 3.68242 Pa/m across the same velocity range, with a similar dip at 0.5 m/s potentially caused by localized flow reorganization. At 8% volume fraction, ΔP/L rises steadily from 0.95156 to 6.55918 Pa/m as velocity increases, reflecting a more linear response indicative of a transition toward mixed suspension flow. Overall, the effect of velocity becomes increasingly significant at higher particle concentrations, marking a clear shift from bed-dominated to suspension-dominated flow regimes.

Figure 5-2 illustrates the variation in pressure drop (ΔP) with hydrate particle volume fraction at three inlet velocities: 0.2, 0.5, and 0.8 m/s, based on a 0.2-second averaging interval. Across all flow rates, a clear nonlinear increase in pressure drop is observed with rising particle volume fraction, consistent with hydrate–water slurry behavior. Power-law trendlines were applied to emphasize this relationship. The trend is in accordance with the results observed by Zhang et al. in their experiments [43], further supporting the validity of the model used in this study.

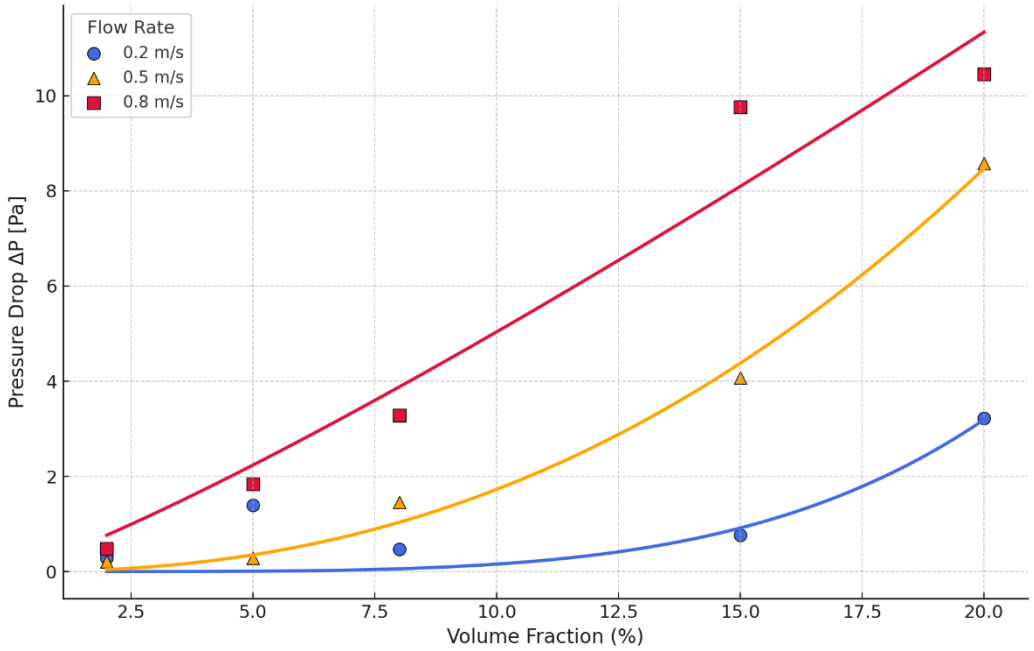


Figure 5-2: Pressure drop vs. volume fraction in horizontal hydrate–water flow.

At the lowest flow rate (0.2 m/s), the increase in pressure drop is relatively moderate, reflecting limited particle entrainment and weaker shear-induced drag. At 0.5 m/s, the pressure drop rises more sharply with increasing volume fraction, particularly beyond 8%, indicating an early transition into a pseudo-slurry regime where wall effects and particle clustering intensify flow resistance. The most pronounced increase occurs at 0.8 m/s, where particles are more effectively suspended, and interphase drag becomes the dominant mechanism of energy loss.

The power-law fitting confirms that the system's hydraulic resistance becomes more sensitive to particle concentration as flow energy increases. These findings reinforce earlier observations about the coupled effects of hydrate loading and velocity on pressure loss in horizontal flow. Furthermore, the consistent shape of the pressure–volume fraction curves across different velocities suggests a scalable and predictable trend, even during the early stages of flow development.

Overall, these results—captured within a 0.2-second simulation window—offer a reliable snapshot of the system’s initial behavior, prior to the onset of long-term effects such as particle migration,

bed stabilization, or outlet losses. This supports the use of truncated time-window analysis as a valid approach for studying early-stage dynamics in transient multiphase flow simulations.

5.3.2. Particle retention and depletion

Figure 5-3 shows the number of particles remaining in the domain over time for different flow rates and volume fractions. Higher flow velocities lead to faster particle depletion, while higher volume fractions result in slower particle exit and longer retention within the domain.

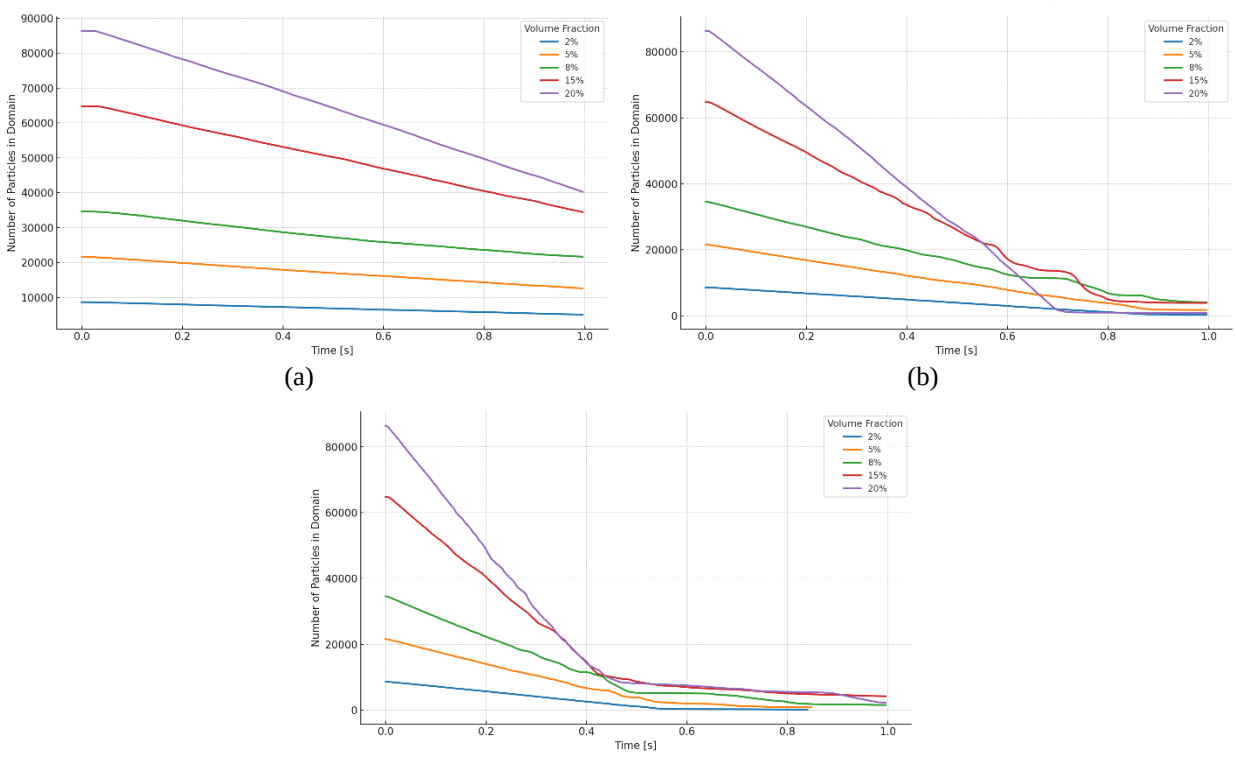


Figure 5-3: Number of particles vs. Time for different hydrate volume fractions at inlet velocities of (a) 0.2 m/s, (b) 0.5 m/s, and (c) 0.8 m/s.

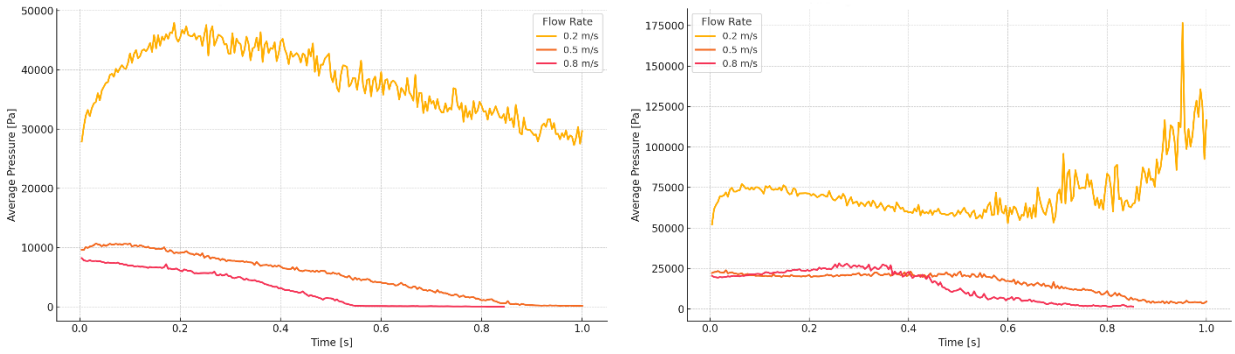
The temporal evolution of hydrate particle count within the domain offers insight into transport dynamics and residence time under varying flow conditions. As shown in Figure 5-3, particle depletion is strongly influenced by inlet velocity and initial volume fraction, governed by the balance of fluid momentum, particle inertia, and drag. At 0.2 m/s, particle movement is slow, with gradual depletion across all volume fractions. Higher concentrations (15% and 20%) retain more

particles, indicating dominant gravitational settling and limited suspension, while lower fractions (2% and 5%) show modest losses due to weak flow momentum and bed formation. At 0.5 m/s, particle exit becomes more noticeable. Dilute cases deplete steadily, whereas higher concentrations retain more particles due to increased particle–particle interactions and bulk resistance. This velocity marks a transition from stratified to partially suspended flow. At 0.8 m/s, rapid entrainment occurs across all volume fractions. Even at high concentrations, strong convective forces overcome settling, leading to consistent particle loss—a behavior typical of suspension-dominated regimes in horizontal slurry transport.

Volume fraction also affects retention: higher concentrations yield greater initial particle counts and longer presence in the domain, driven by enhanced collisions, drag buffering, and increased mixture viscosity. These trends are well captured by the CFD–DEM model and the Gidaspow's drag model, highlighting the method's effectiveness in resolving fluid–solid and solid–solid interactions.

5.3.3. Pressure evolution

Figure 5-4 shows the evolution of average pressure over time for five hydrate volume fractions (2%, 5%, 8%, 15%, and 20%). Each case starts with the same number of particles, allowing a clear comparison of how flow velocity affects pressure and particle movement. The results show that pressure behavior in hydrate–water flows is strongly influenced by velocity, with differences mainly caused by how quickly particles move through and exit the domain.



(a) (b)

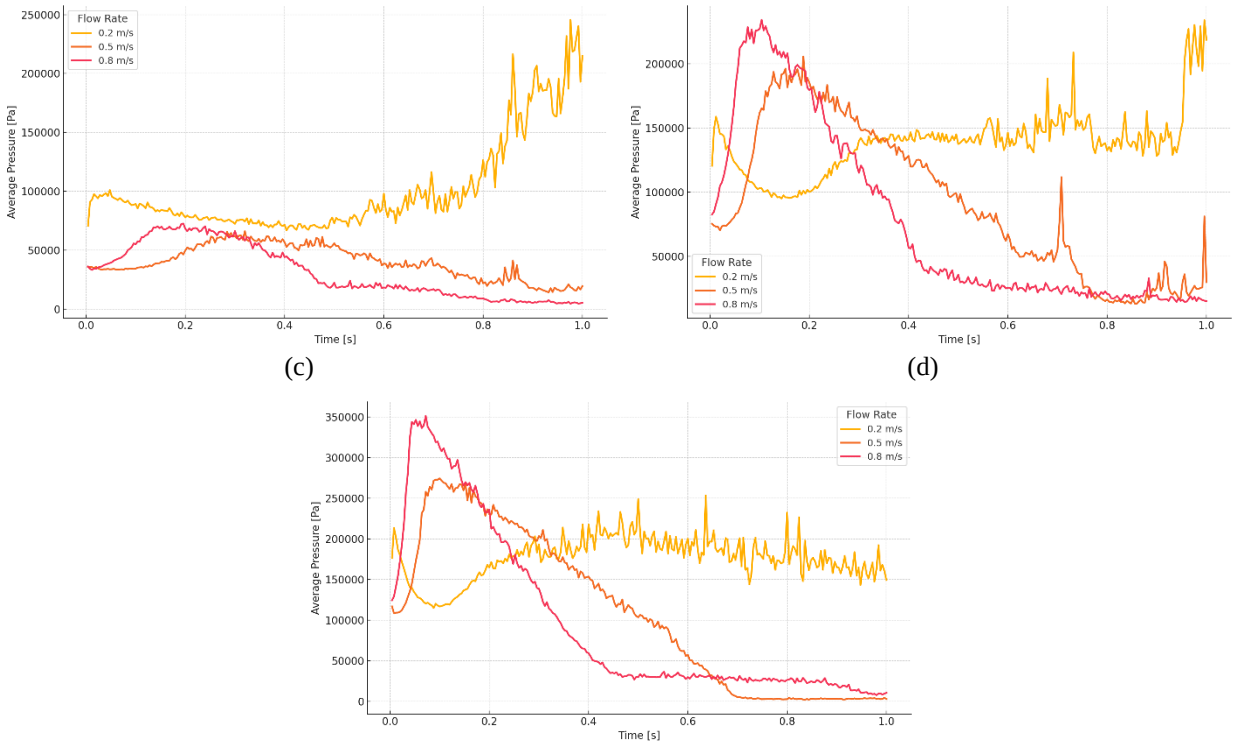


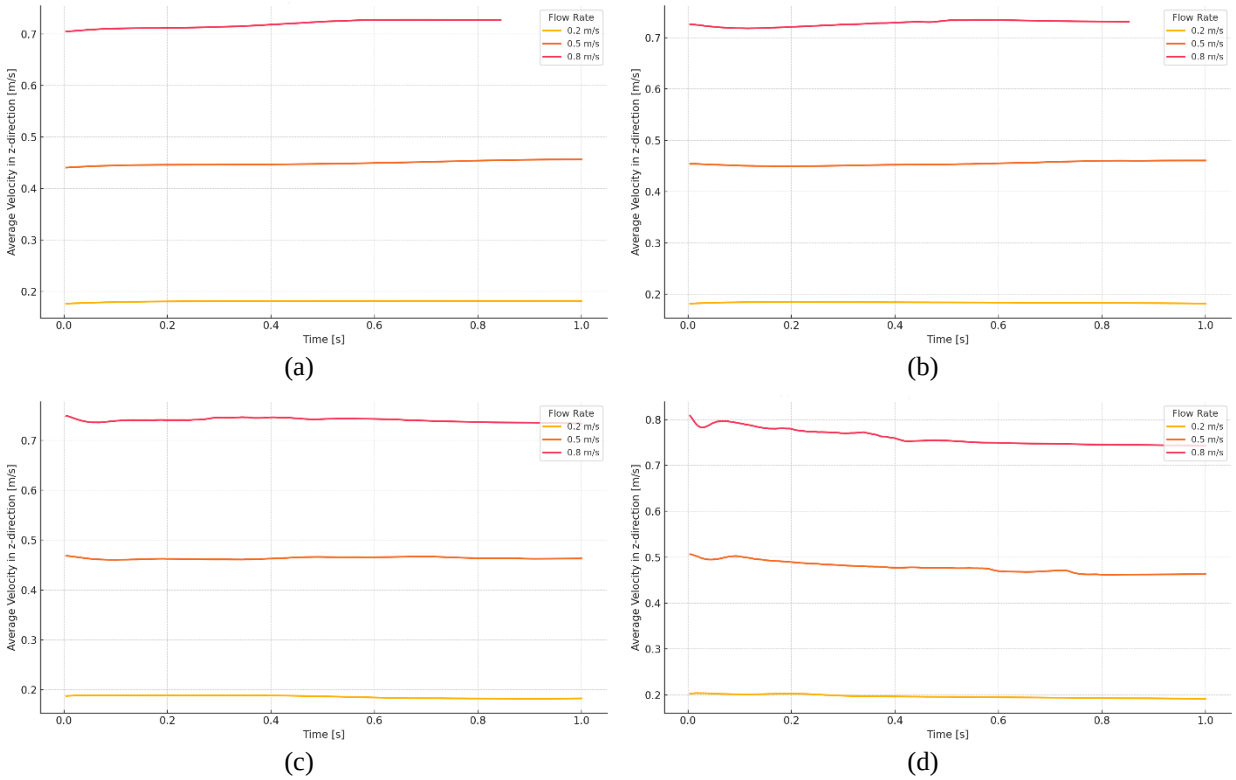
Figure 5-4: Figure 3. Average pressure vs. time for hydrate–water flow at five volume fractions ((a) 2%, (b) 5%, (c) 8%, (d) 15%, and (e) 20%) with three inlet velocities (0.2, 0.5, and 0.8 m/s).

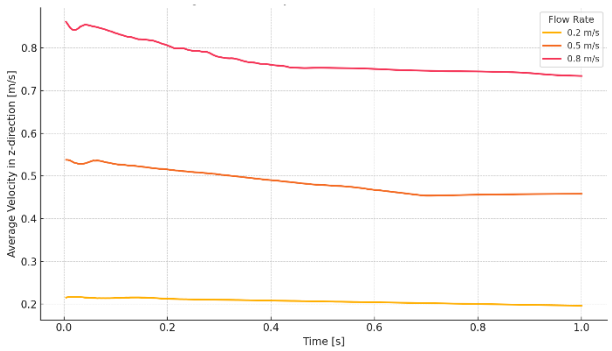
At 0.2 m/s, pressure builds gradually and remains fairly stable, especially at lower volume fractions. This reflects a bed-dominated regime, where particles settle and remain in the pipe, maintaining drag. At higher velocities (0.5 and 0.8 m/s), pressure rises faster but drops sooner, as particles are carried out more quickly, reducing solid loading and flow resistance. These trends match known multiphase flow behavior. Higher velocities apply more force to the particles, helping to lift and move them. Early pressure peaks are due to drag from suspended particles, while later drops reflect particle loss. At higher concentrations (15% and 20%), pressure rises are sharper due to denser packing and more interactions, but also fall off quickly at high speeds. This highlights how concentration and particle retention affect resistance. The results support using early time windows in pressure analysis. After many particles leave (around 0.6–0.8 seconds), pressure no longer reflects the initial flow state. Focusing on early stages gives a clearer picture of system behavior.

Overall, increasing flow rate at fixed particle concentration leads to faster movement, quicker particle loss, and earlier pressure drop. The strong link between pressure and particle behavior confirms the CFD–DEM model's effectiveness in capturing hydrate transport in horizontal flows.

5.3.4. Fluid velocity trends

To further explore the impact of hydrate particles on flow behavior, the spatially averaged axial velocity (z-direction) was evaluated over time for all simulation cases. Figure 5-5 shows how the average velocity changes at different inlet flow rates for each hydrate volume fraction. Each profile represents the average velocity across all cells in the domain, capturing how the flow evolves as particles move through and leave the system.





(e)

Figure 5-5: Average axial velocity (z-direction) vs. time for hydrate–water flow at fixed volume fractions (a) 2%, b) 5%, c) 8%, d) 15%, and e) 20%) with varying inlet velocities (0.2, 0.5, and 0.8 m/s).

These velocity patterns help explain how hydrate–water flows behave over time. They show an early slowdown due to drag from particles, followed by a gradual increase in velocity as particles leave the domain. These results match well with the earlier trends seen in pressure and particle count. At low concentrations (2%), velocity stays mostly steady for all flow rates, with only a small initial drop due to drag. At medium concentrations (5–8%), especially at higher velocities, there is a more noticeable slowdown, followed by a gradual increase as fewer particles remain and resistance goes down. In dense flows (15–20%), velocity drops quickly at the start, especially at 0.8 m/s, showing strong drag and energy loss. As particles are carried out of the pipe, the flow slowly speeds up again. This behavior is similar to what was seen in the pressure results, where pressure first rises and then falls for the same reasons.

Overall, these findings support a time-based view of hydrate transport. At first, the particles slow the flow, but as they leave, the system starts to recover. This also supports using shorter time frames when analyzing pressure and velocity, so the results are not affected by late-stage particle loss.

5.3.5. Drag force behavior

To better understand the interaction between fluid and particles in hydrate–water flow, the average fluid–particle drag force was tracked over time. This force shows how strongly the fluid pushes on the hydrate particles and depends on both particle concentration and flow rate. Figure 5-6 presents

how the drag force changes over time for five volume fractions (2%, 5%, 8%, 15%, and 20%) at three different inlet velocities. This analysis provides a direct look at momentum exchange between the fluid and particles and supports earlier findings related to pressure, velocity, and particle count.

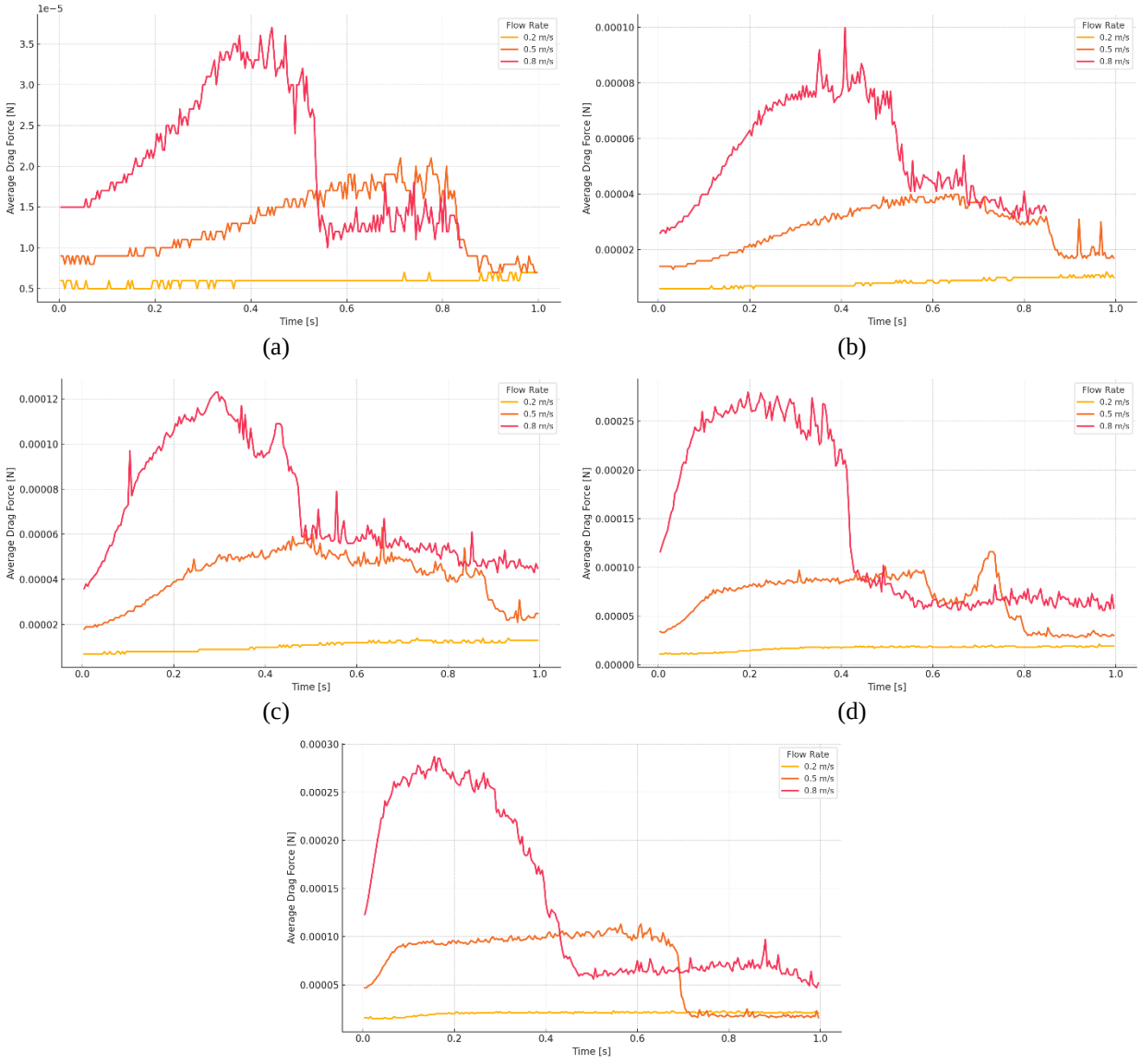


Figure 5-6: Average fluid–particle drag force vs. time for hydrate–water flow at fixed volume fractions (a) 2%, b) 5%, c) 8%, d) 15%, and e) 20%) and varying inlet velocities (0.2, 0.5, and 0.8 m/s).

The drag force plots show how fluid–particle interaction changes as flow conditions evolve. This force reflects the resistance created by particles in the flow and changes based on how many

particles are present, how fast the flow is, and how quickly particles leave the domain. At 2% volume fraction, drag remains low and steady, with only a small peak at the start. This low-drag behavior is typical of dilute systems where particles are quickly carried out, resulting in little ongoing resistance. The pattern is similar to the steady pressure and velocity trends and fast particle loss shown in earlier figures. At 5% and 8%, drag becomes more noticeable, especially at 0.8 m/s. Early peaks show a short period of strong resistance before the drag drops again as particles leave the system. These cases reflect a transition zone, where flow starts to suspend particles more, but not for long. At 15% and 20%, drag force is much stronger in the beginning, especially at higher flow rates. The sharp peaks show intense interaction between the fluid and the densely packed particles. As particles move out of the pipe, drag steadily decreases. This matches the high pressure at the start, early velocity drops, and gradual particle removal seen earlier.

Overall, drag force is highest when both the flow rate and particle concentration are high. As particles leave, the force drops, showing how flow resistance weakens over time. These results clearly show that hydrate transport changes with time, starting with strong resistance and followed by recovery. This also supports using short time frames for analyzing pressure and velocity, since later results may be affected by particle loss.

5.3.6. Interparticle force dynamics

To examine the mechanical interactions within the solid phase of hydrate–water flow, the average magnitude of interparticle contact forces was tracked over time for each case. These forces, calculated using the Hertz–Mindlin contact model with the cohesive SJKR extension, account for both normal and tangential contact components, including friction and cohesion. Measured in Newtons (N), they reflect how hydrate particles respond to fluid momentum, crowding, and dynamic loading. Figure 5-7 shows how these forces change at different volume fractions, providing insight into particle clustering, mechanical resistance, and energy dissipation during the simulation.

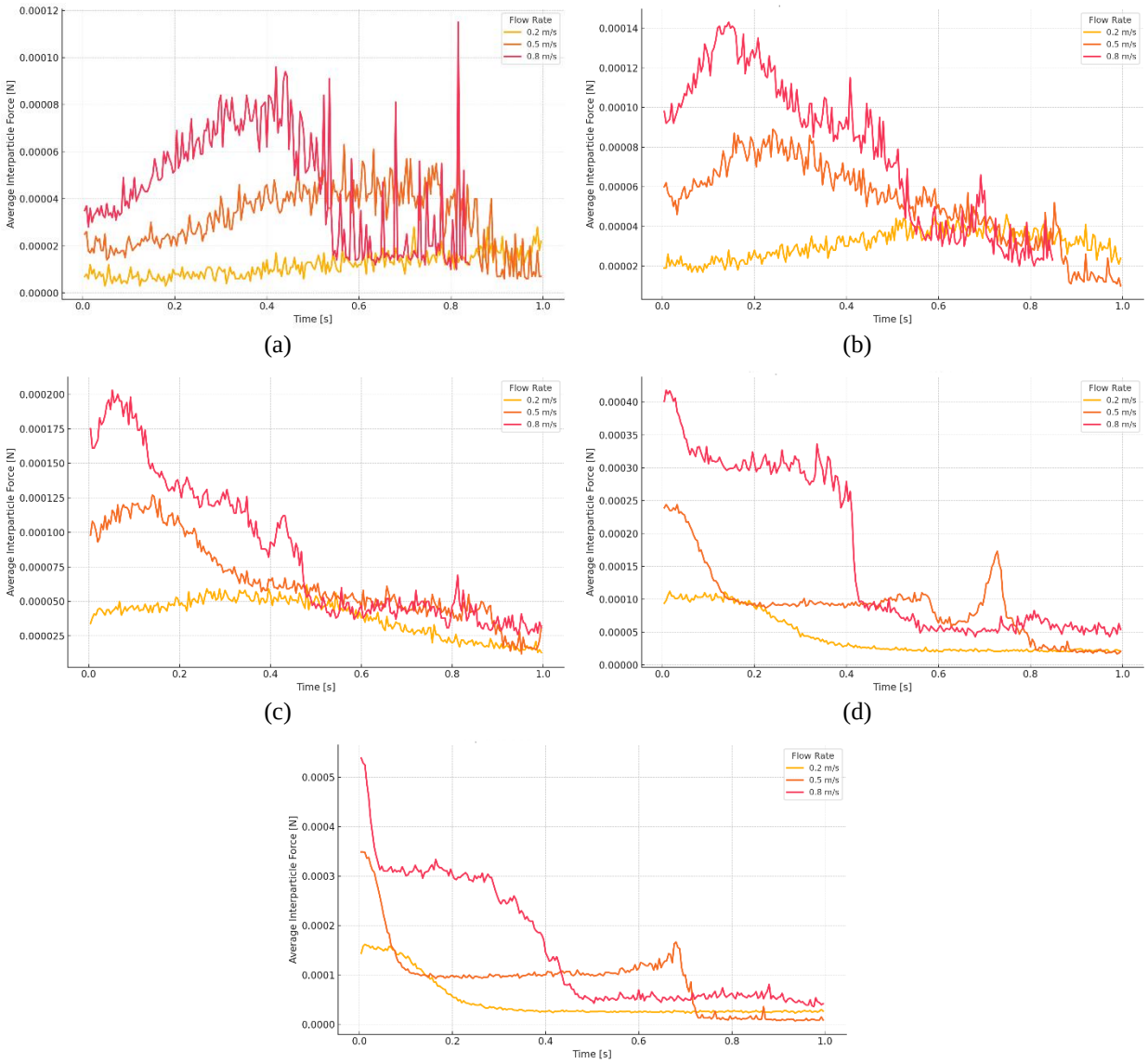


Figure 5-7: Average interparticle contact force vs. time at fixed hydrate volume fractions (a) 2%, b) 5%, c) 8%, d) 15%, and e) 20%), evaluated for three inlet velocities (0.2, 0.5, and 0.8 m/s).

Each curve illustrates how interparticle collisions and cohesive forces evolve over time, depending on both particle concentration and flow rate. These trends help explain how the solid phase behaves and complement the earlier fluid-phase observations. At 2% volume fraction, contact forces are very low. A small peak appears only at 0.8 m/s, followed by a quick drop, indicating that particles interact briefly before being carried out of the domain with minimal sustained contact. At 5% and 8%, the flow enters a transitional state. Noticeable early peaks, especially at high flow rates, reflect moderate crowding and collision activity. These forces gradually decline as particles leave the

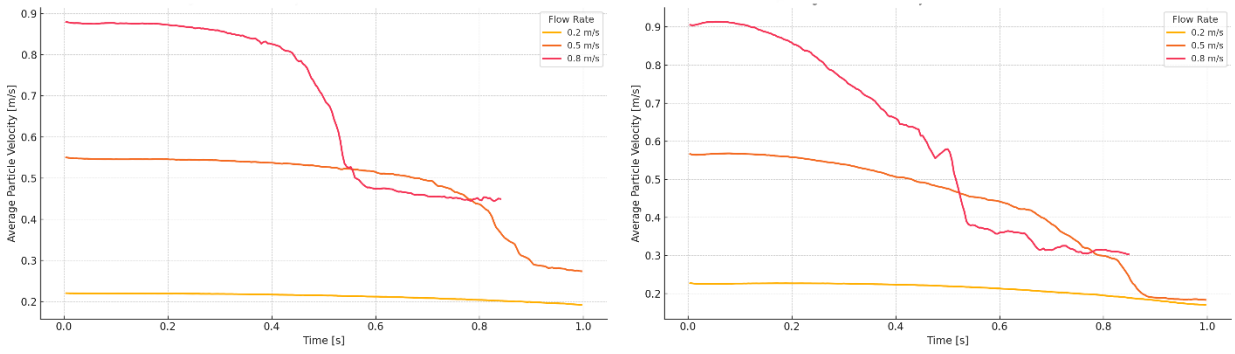
pipe, matching the earlier drop in drag, pressure, and the rise in velocity. This shows that as mechanical resistance fades, the flow becomes more continuous. The 15% and 20% cases show the highest contact forces, especially at 0.8 m/s. Strong early peaks reflect intense particle clustering, energy loss through collisions, and increased resistance. Over time, these forces decrease as particles exit the domain, following similar trends seen in velocity recovery and pressure drop.

Overall, contact forces are strongest when both particle concentration and velocity are high. As particles are removed, these forces decrease in a predictable way. This matches the earlier trends in drag, pressure, and velocity, showing that early flow resistance is mainly caused by particle crowding, while later recovery happens as the system becomes more dilute.

These interparticle forces clearly indicate the presence of mechanical clustering and cohesion, and show that the SJKR model effectively captures this behavior. Alongside drag, pressure, and velocity data, these results support using shorter time windows for analysis, to better capture early flow resistance without distortion from later particle loss. This approach improves how both fluid and particle interactions are interpreted during hydrate transport.

5.3.7. Particle velocity trends

To complement the analysis of fluid behavior and interphase forces, the average velocity of hydrate particles was tracked over time for each simulation case. This measure shows how the solid phase moves in response to different inlet velocities and volume fractions. It captures how particles are carried by the fluid, interact with each other, and gradually leave the domain. Figure 5-8 presents the time evolution of particle velocity for five volume fractions (2–20%) at three flow rates.



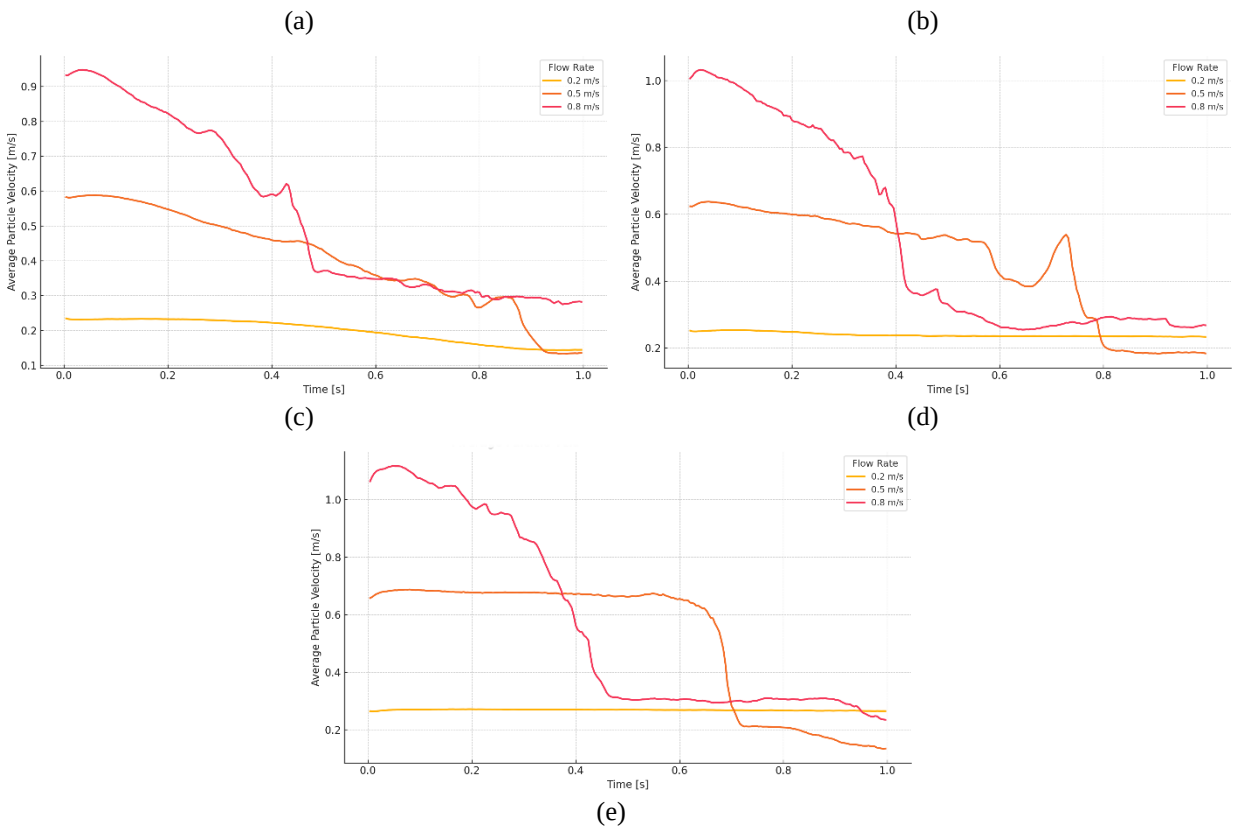


Figure 5-8: Average particle velocity vs. time for hydrate–water flow at fixed volume fractions (a) 2%, b) 5%, c) 8%, d) 15%, and e) 20%) and inlet velocities of 0.2, 0.5, and 0.8 m/s.

These trends give a clear view of particle motion and how it changes under the effects of fluid drag, collisions, and particle removal. The results match earlier findings on pressure, drag force, fluid velocity, and particle interactions, showing how movement is shaped by both fluid forces and particle buildup. At 2% volume fraction, particle speed stays high across all flow rates, showing that particles are easily moved by the flow and experience little collision. This behavior is typical of dilute flows, where particles move freely and are quickly transported. It reflects earlier observations of low drag, weak contact forces, and steady pressure. For 5% and 8% cases, particle velocity rises quickly at the beginning, especially at higher flow rates like 0.8 m/s. This initial increase shows that the flow strongly pushes the particles. Over time, particle speed drops as they leave the domain and interactions lessen. These trends follow similar patterns in drag and pressure—starting strong, then slowing as particles are removed. At higher concentrations (15% and 20%), particles move more slowly at first because they are packed closer together and stick more. But at higher flow speeds, particles begin to move faster later in the simulation, as the

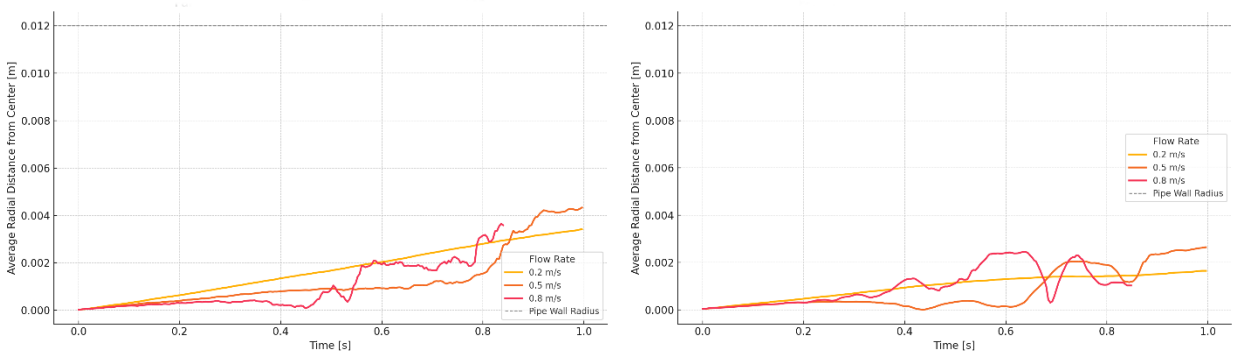
crowding breaks apart and resistance goes down. This shift from blocked movement to free motion matches earlier drops in drag and rises in fluid velocity.

Overall, the particle velocity patterns support what has been seen throughout the study: high concentrations and flow speeds cause strong resistance early on, followed by a gradual return to smoother flow as particles leave. These results show the strength of the CFD–DEM model in capturing hydrate transport.

This analysis also supports using truncated time windows to focus on the most meaningful simulation period. After 0.6 to 0.8 seconds, particle loss begins to dominate and may distort average values. By focusing on early to mid-stage dynamics, the analysis stays grounded in the physical behavior that defines hydrate–fluid interactions.

5.3.8. Radial distribution of particles

To understand how hydrate particles spread within the pipe cross-section during transport, the average radial distance from the pipe center was calculated over time. This value, based on the mean x and y positions of the particles at each time step, shows whether particles stay near the center or move toward the walls. The radial distance was computed as $R = \sqrt{\text{ave}(x)^2 + \text{ave}(y)^2}$ and compared to the pipe radius (0.012 m) to assess how much particles spread out or interact with the wall. Figure 5-9 shows how particle positions change over time for different volume fractions and inlet velocities, reflecting the effects of flow rate, concentration, and particle interactions.



(a) (b)

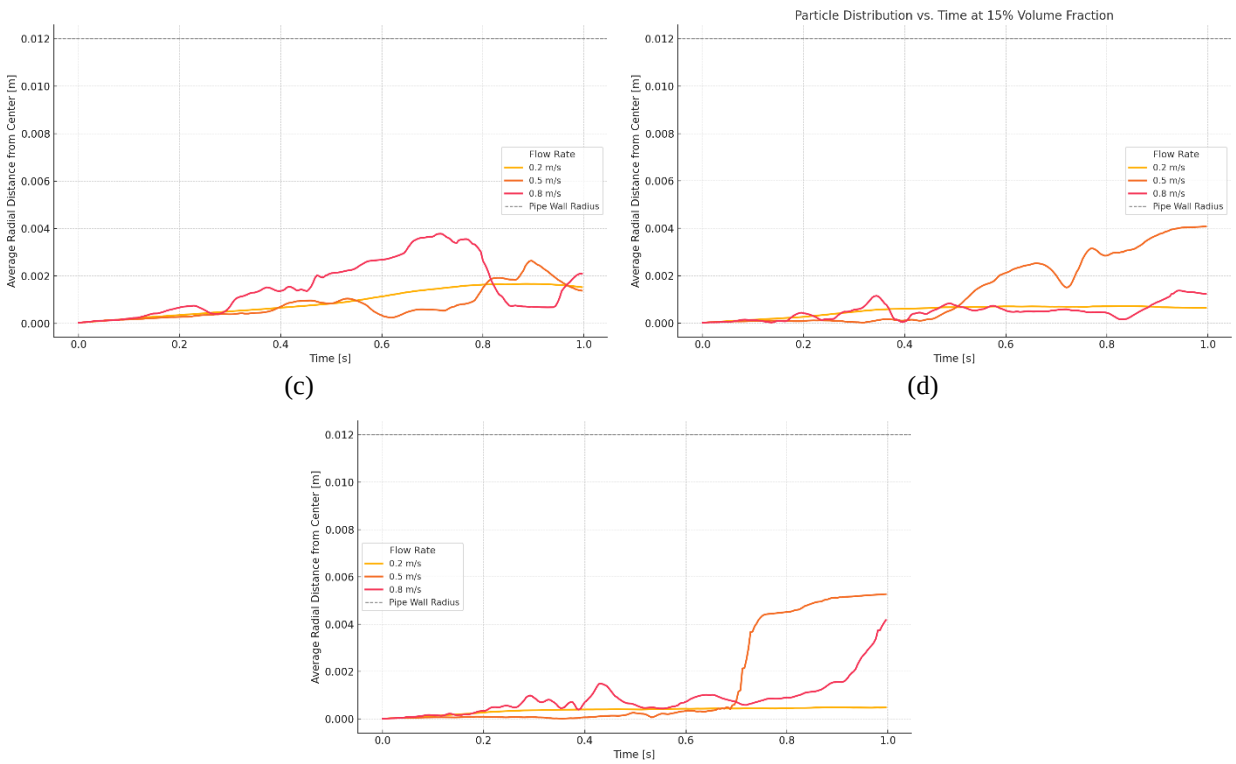


Figure 5-9: Average radial distance from the pipe center for hydrate particles over time, shown for five volume fractions (a) 2%, b) 5%, c) 8%, d) 15%, and e) 20%) at three inlet velocities (0.2, 0.5, and 0.8 m/s).

These results offer a clear view of how particles are distributed in the pipe and how that changes with flow conditions. At a 2% volume fraction, particles stay close to the center throughout the simulation, with little variation across flow rates. This suggests a well-mixed suspension with minimal spreading or contact between particles, which matches the low drag, low interparticle forces, and steady particle speeds seen under these conditions. At 5% and 8%, flow rate starts to have more influence. At higher velocities, particles shift slightly outward, showing mild spreading caused by stronger fluid forces. However, they still remain generally centered, indicating that moderate loading does not yet lead to significant movement toward the pipe walls. For 15% and 20% volume fractions, particles move farther from the center, especially at 0.8 m/s. The average distance approaches, but does not exceed, the pipe wall radius. This pattern suggests that particles begin to gather near the walls due to stronger collisions and higher shear. It also matches early peaks in drag, interparticle forces, and pressure, along with the slower particle speeds seen early

in the flow. As particles leave the system, the average radial position either decreases or levels off, showing a return to a more centered suspension.

These results show that particle flow behavior strongly depends on concentration. Lower concentrations result in more uniform, center-aligned flow, while higher concentrations, particularly at faster flow rates, lead to temporary buildup near the walls. This redistribution affects and reflects the changes observed in drag, pressure, flow velocity, and collisions.

As with other findings, this analysis supports focusing on the first 0.6 to 0.8 seconds of the simulation. After this point, particle loss dominates the results and may hide the actual transport behavior. Studying the early phase allows for a clearer understanding of how hydrate particles interact with the fluid and with each other during active flow.

5.3.9. Statistical overview of particle-phase behavior

To complement the graphical analysis and highlight key numerical trends, a summary table was created using time-averaged values from the first 0.6 seconds of each simulation. This early phase captures the most meaningful part of hydrate–water transport, before significant particle loss occurs. It reflects how flow rate, particle concentration, and interphase interactions shape system behavior. Table 5-4 reports average values for drag force, interparticle force, particle speed, and radial distance from the pipe center, offering a clear overview of particle dynamics across all cases.

Table 5-4: Truncated (0–0.6 s) time-averaged metrics for hydrate–water flows at varying flow rates and volume fractions.

No.	Particle Velocity (m/s)	Volume Fraction (%)	Drag (N)	Interparticle Force (N)	Particle Velocity (m/s)	Radial Distance
1	0.2	2	0.00001	0.00001	0.21784	0.00100
2	0.2	5	0.00001	0.00003	0.22387	0.00070
3	0.2	8	0.00001	0.00005	0.22337	0.00053
4	0.2	15	0.00002	0.00006	0.24308	0.00042
5	0.2	20	0.00002	0.00006	0.27027	0.00030
6	0.5	2	0.00001	0.00003	0.53880	0.00057
7	0.5	5	0.00003	0.00007	0.52488	0.00023
8	0.5	8	0.00004	0.00008	0.49853	0.00048
9	0.5	15	0.00008	0.00012	0.57220	0.00035
10	0.5	20	0.00009	0.00012	0.67469	0.00009
11	0.8	2	0.00002	0.00005	0.79723	0.00045
12	0.8	5	0.00006	0.00010	0.72220	0.00076
13	0.8	8	0.00009	0.00012	0.68075	0.00109

No.	Particle Velocity (m/s)	Volume Fraction (%)	Drag (N)	Interparticle Force (N)	Particle Velocity (m/s)	Radial Distance
14	0.8	15	0.00019	0.00024	0.69520	0.00039
15	0.8	20	0.00018	0.00022	0.75325	0.00051

This brief snapshot reveals several important patterns. Both drag force and interparticle force increase steadily with volume fraction at all flow rates. For instance, at 0.2 m/s, the drag force rises from about 0.00136 N at 2% concentration to over 0.00442 N at 20%. The interparticle force also grows from roughly 0.00062 N to 0.00449 N across the same range. This reflects how denser particle suspensions lead to greater fluid resistance and more frequent collisions. Interestingly, particle speed does not always decrease with higher concentrations. At 0.2 m/s, average particle velocity increases from about 0.21741 m/s at 2% to 0.27084 m/s at 20%. This suggests that once particle clusters begin to break apart, stronger fluid drag may help maintain or even increase particle movement, possibly due to tighter alignment with the flow and reduced dispersion. Radial distance trends reveal spatial effects. At low volume fraction and 0.2 m/s, particles are more spread out from the center (average distance ≈ 0.00108 m), while at 20% concentration, they are closer to the center (≈ 0.00037 m). This suggests that dense suspensions may form temporary core-like structures near the centerline before gradually spreading outward, matching the behavior seen in the radial distribution plots.

Overall, Table 5-4 confirms and quantifies the key observations from the figures. Systems with higher concentrations and faster flow show stronger early resistance, greater mechanical interaction, and more structured particle behavior. These numerical results support the interpretation of flow regime changes, clustering, and particle movement in hydrate–water systems.

5.3.10. Velocity-based visualization of particle distribution: temporal and concentration effects

To clarify the dynamic behavior of hydrate particles in water-dominated pipe flow, two sets of cross-sectional visualizations were generated based on particle velocity magnitude. These plots provide a detailed qualitative assessment of flow structure and particle distribution under varying

conditions. Figure 5-10 illustrates the temporal evolution of particle behavior at a fixed volume fraction, while Figure 5-11 captures the effect of concentration at a fixed time. In both cases, the simulations were conducted at an inlet velocity of 0.5 m/s—representative of an intermediate flow regime where both suspension and stratification phenomena are observed.

Each image was rendered by projecting the simulation domain from the Z+ direction, thus presenting a planar cross-section of the pipe inlet. Particle velocities are represented via a continuous color map ranging from blue (0 m/s) to yellow (0.7 m/s), enabling a direct visual interpretation of kinetic energy distribution and interphase momentum exchange.

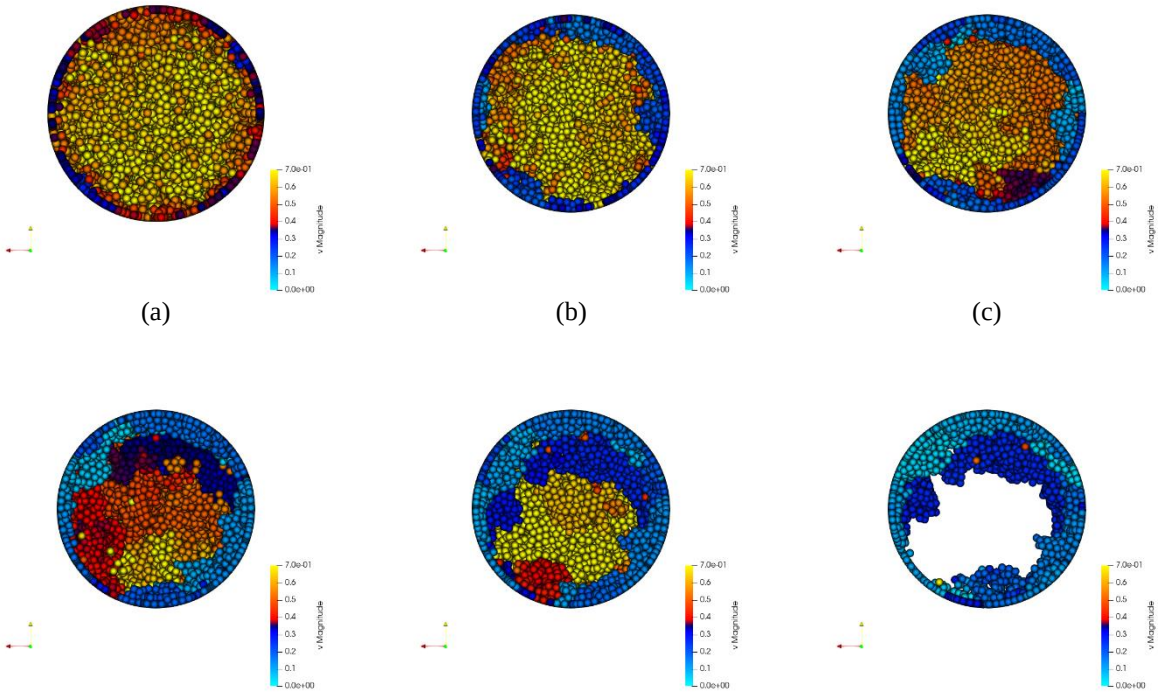


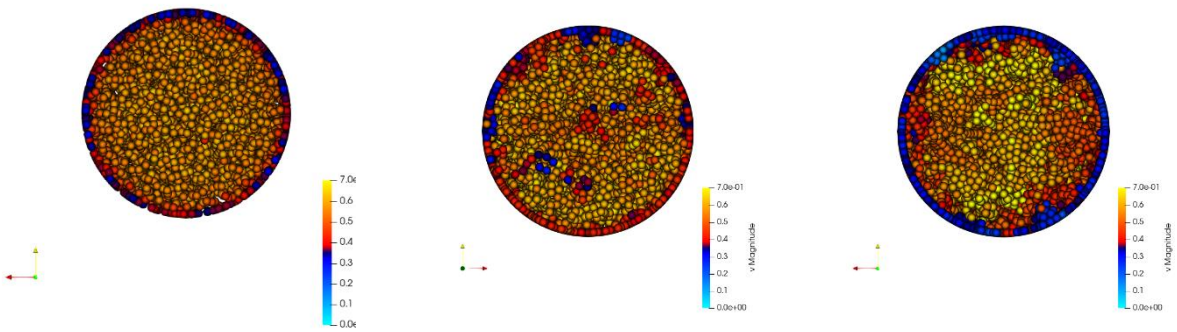
Figure 5-10: 2D cross-sectional velocity plots at the pipe inlet (z = 0 m) for hydrate–water flow at a volume fraction of 15% and velocity of 0.5 m/s. Snapshots are shown at time intervals of a) 0.08 s, b) 0.24 s, c) 0.56 s, d) 0.72 s, e) 0.88 s, and f) 0.96 s.

As time progresses, significant changes are observed in both the suspension quality and velocity distribution of the hydrate particles. During the early stages (0.08–0.24 s), the particles are largely well dispersed, characterized by high velocity magnitudes and limited interaction with the pipe walls. This behavior indicates strong momentum coupling with the fluid phase and minimal

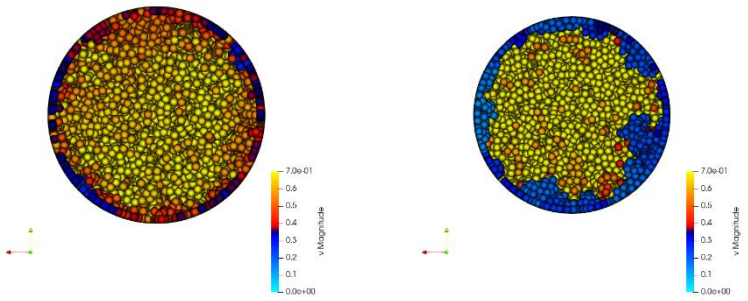
mechanical resistance within the particle assembly. Beyond 0.56 s, heterogeneity in particle velocity becomes increasingly apparent. Localized regions of lower velocity, particularly near the wall and within emerging particle clusters, begin to form. This behavior reflects the development of cohesive interactions and enhanced frictional contacts between particles. Over time, particles tend to migrate toward the pipe boundaries, where cohesive and mechanical forces contribute to the formation of semi-stable, low-mobility aggregates.

The increasing clustering and near-wall retention over time emphasize the importance of early hydrodynamic resistance, followed by progressive particle loss and reorganization. This behavior is strongly governed by the balance between fluid momentum, cohesive forces, and crowding effects. In the later stages (0.88–0.96 s), a depletion of particles in the central region becomes visible, indicating that many particles have exited the domain. Although the observations are made at the inlet, the depletion reflects the continuous downstream transport and domain flushing processes. These trends underscore a progressive transition from an initially uniform suspension to a partially stratified flow, where mechanical interactions and reduced flow energy promote clustering and wall-layer formation.

The time of 0.24 seconds was selected to capture a developed yet representative phase of hydrate–water transport. By this stage, particle–fluid interactions and clustering behavior have clearly emerged, while significant particle depletion has not yet occurred. This ensures that suspension quality, wall interactions, and flow structure can be meaningfully compared across different volume fractions under consistent hydrodynamic conditions.



(a) (b) (c)



(d) (e)

Figure 5-11: Cross-sectional inlet views of hydrate particle velocity magnitude at a fixed simulation time of 0.24 s and a flow velocity of 0.5 m/s. Volume fractions considered are a) 2%, b) 5%, c) 8%, d) 15%, and e) 20%.

The influence of particle concentration on hydrate transport is clearly demonstrated through comparative analysis of the five cases. At low concentrations (2–5%), particles are relatively uniformly distributed with moderate to high velocities across the domain. The flow is dominated by fluid drag forces, and cohesive or frictional interactions between particles are minimal. Suspension quality remains high, and wall interaction is negligible. At intermediate concentrations (8%), early clustering phenomena become evident. Slight velocity gradients emerge, with regions of reduced velocity near the pipe walls, indicating the onset of mechanical crowding and shear-induced migration. Although the bulk of the particles are still suspended, interparticle interactions begin to affect the flow structure. At higher concentrations (15–20%), substantial changes in particle distribution are observed. Dense packing occurs near the pipe walls, where lower particle velocities are prevalent, while the central region remains partially mobile. These trends indicate intensified interparticle cohesion, frictional resistance, and localized energy dissipation. The visual progression from predominantly yellow regions in dilute flows to predominantly blue and orange regions in dense flows captures the transition toward a pseudo-slurry regime, where the solid phase exhibits collective behavior characteristic of semi-consolidated flows.

Overall, these results are consistent with previous quantitative analyses of drag force, interparticle contact force, and pressure drop. Increasing particle concentration promotes greater mechanical resistance, decreases suspension efficiency, and fosters spatial heterogeneity within the flow. The

visualizations reinforce the critical role of concentration in governing hydrate transport behavior and validate the CFD–DEM framework for capturing complex multiphase flow dynamics.

5.4. Conclusions

This study employed a two-way coupled CFD–DEM framework to systematically investigate the dynamic behavior of hydrate particles in water-dominated horizontal pipe flows. A range of flow rates and hydrate volume fractions were analyzed to capture the interplay between fluid momentum, interphase forces, particle interactions, and spatial distributions during the early stages of flow development.

The results revealed that both particle concentration and flow velocity significantly influence hydrate transport characteristics. Higher volume fractions led to greater pressure drops, stronger interparticle forces, and enhanced mechanical resistance, while increased flow velocities promoted particle suspension, accelerated depletion, and reduced radial dispersion. Early-time analysis demonstrated that hydrate–water flows transition from initially high resistance due to clustering and wall interactions to progressively more dilute, free-moving regimes as particles exit the domain.

Velocity-based visualizations further highlighted key transitions in particle behavior, with the emergence of clustering, near-wall accumulation, and heterogeneous flow structures becoming more pronounced at higher concentrations and longer times. Quantitative metrics such as drag force, interparticle force, particle velocity, and radial distribution consistently validated the time-resolved interpretation of hydrate transport dynamics.

Overall, the findings emphasize the critical role of hydrate concentration, fluid energy, and interphase coupling in governing flow assurance risks in multiphase pipeline systems. The study also highlights the value of truncated time window analysis for accurately capturing early-stage dynamics before depletion-driven artifacts dominate. These insights contribute to a deeper understanding of hydrate slurry transport and offer a foundation for improving predictive models and mitigation strategies in flow assurance applications.

Chapter 6: Conclusions and recommendations

6.1. Summary of key findings

This study applies complementary approaches, including data-driven machine learning, thermodynamic modeling, and CFD–DEM simulations, to improve the prediction and management of gas hydrate systems. The key scientific findings and conclusions derived from the results are summarized below.

The machine learning-based models developed in Chapter 2 demonstrate that methane hydrate equilibrium temperatures can be accurately predicted across a wide range of saline environments. The results show that using ion mole fractions as input features allows the models to generalize beyond the original training data, making it possible to predict hydrate formation conditions for up to 36 different single-salt systems based on six cations and six anions. Among the models tested, the Extremely Randomized Trees (ET) model achieved the highest accuracy, with an RMSE of 0.6248 K for the main dataset and 0.3520 K for the independent $MgBr_2$ system. These results demonstrate that machine learning models can match or outperform existing predictive correlations while offering wider chemical coverage. Importantly, the findings highlight that the ET model not only accurately interpolates within known systems but also extrapolates reliably to new ionic compositions. This supports the use of machine learning as a practical tool for hydrate risk management and process design in natural gas storage, transportation, and desalination technologies.

In Chapter 3, extending the modeling to methane hydrates involving water-soluble hydrate formers revealed that complex molecular-level descriptors, such as polar surface area and rotatable bond counts, are critical to accurately predicting equilibrium conditions. The results demonstrate that Random Forest (RF) models, when combined with careful feature selection and a novel former-based data-splitting approach, can achieve high generalization performance (R^2 = 0.930, RMSE = 1.71 K, AARD = 0.48%). The use of Shapley Additive Explanations (SHAP) shows that even features initially perceived as less important have a meaningful impact on model stability, especially in distance-based algorithms like SVM and MLP. These findings emphasize that hydrate former selection should consider not only simple physicochemical properties but also molecular

flexibility and polarizability. The work advances the interpretability of machine learning in hydrate science and provides a framework for rational selection and design of hydrate formers for targeted gas storage and separation applications.

Chapter 4 explores gas hydrate dissociation behavior under confined conditions. The results show that dissociation temperatures shift to lower values with decreasing pore size, in agreement with the Gibbs–Thomson effect, and that larger pores tend to behave like bulk systems. The thermodynamic activity model developed achieves acceptable predictive accuracy, confirming that incorporating activity coefficients and confinement effects into thermodynamic modeling is essential for confined hydrate systems. Machine learning models further validate that critical temperature, system pressure, and pore size are the dominant predictors for hydrate stability under confinement. These results demonstrate that simple scaling laws, such as those based solely on pore size, are insufficient to describe hydrate behavior in complex porous networks. Real hydrate stability under confinement is influenced by many factors, including capillary pressure, interfacial effects, and gas properties, which cannot be captured by simple formulas. This highlights the necessity of applying both physics-based and data-driven approaches to achieve accurate predictions of hydrate stability under confinement. This insight has important implications for understanding hydrate behavior in natural reservoirs, carbon sequestration sites, and engineered porous media.

In Chapter 5, the CFD–DEM simulations of hydrate slurry flow reveal critical dynamic behaviors that affect flow assurance strategies. The results show that increasing hydrate volume fraction significantly enhances interphase momentum exchange and mechanical resistance, leading to higher pressure drops and earlier clustering of particles near the pipe walls. Cross-sectional velocity profiles confirm that clustering, stratification, and particle-wall adhesion are driven by cohesive and frictional forces between hydrate particles, particularly at high concentrations. Higher inlet velocities help maintain particle suspension initially but accelerate domain heterogeneity and depletion over time. These findings demonstrate that hydrate slurry flow is highly sensitive to early-time dynamics, and that early clustering is a key precursor to flow blockage risks. The results highlight the need for real-time monitoring and dynamic flow control during subsea pipeline operations to prevent hydrate plug formation.

Overall, the integrated results of this thesis show that machine learning models, when properly designed and interpreted, can provide highly accurate and generalizable predictions for hydrate formation and dissociation. Thermodynamic modeling remains essential in confined systems but can be effectively complemented by machine learning. CFD–DEM simulations provide a powerful tool to reveal hydrate particle dynamics that are otherwise impossible to capture experimentally. Overall, the results emphasize that predictive hydrate management requires a multi-scale approach, connecting molecular-level thermodynamics, macroscopic phase behavior, and mesoscopic transport phenomena to ensure safe and efficient energy operations involving gas hydrates.

6.2. Recommendations for future research

This thesis has demonstrated the significant potential of machine learning, thermodynamic modeling, and CFD–DEM simulations in advancing the understanding and predictive capabilities for gas hydrate systems. Nevertheless, several avenues remain open where further research could expand upon these findings and address current limitations.

First, the machine learning models developed for methane hydrate formation conditions can be extended to other gas systems, such as carbon dioxide (CO_2), nitrogen (N_2), and hydrogen sulfide (H_2S). Including hydrate equilibrium data for a wider range of gases would broaden the models' applicability to various real-world scenarios, including carbon capture and hydrogen storage projects. Future research could focus on expanding experimental databases, training generalized models for multi-gas systems, and studying multi-component gas hydrates where interactions between different gases influence phase behavior.

Additionally, while the present models effectively predict formation temperatures for pure water and selected brines, environmental factors such as varying salinity, pH, and the presence of inhibitors were not fully incorporated. Future studies should aim to integrate these variables to enhance model robustness under real field conditions, especially for seawater systems where ionic complexity is higher. Investigating model performance under extreme temperature and pressure conditions relevant to offshore and deep-sea environments would also be valuable.

Regarding the machine learning approach developed for predicting the equilibrium conditions of methane hydrates with water-soluble promoters, future research could focus on extending these models to other gas systems such as carbon dioxide (CO_2), nitrogen (N_2), hydrogen sulfide (H_2S), and their mixtures. This would support the development of more generalizable models applicable to a wider range of hydrate-forming conditions. Additionally, enriching the molecular descriptor space with properties like dipole moment, polarizability, or using 3D and graph-based features from cheminformatics could improve model performance and physical relevance. Expanding the dataset to include more hydrate formers, broader concentration ranges, and incorporating experimental uncertainties or replicate measurements would further strengthen the robustness and reliability of the predictive framework.

In terms of hydrate dissociation under confinement, the results show that both thermodynamic and machine learning models perform well in predicting equilibrium temperatures. However, future research could focus on expanding datasets to include a wider range of porous media, such as natural sediments, coal seams, and carbonate reservoirs. Further, incorporating detailed pore structure characteristics, such as surface roughness, connectivity, and wettability effects, would enable more accurate modeling of hydrate stability under realistic geological conditions. Dynamic modeling approaches that account for changing temperature gradients, fluid flow, and hydrate growth or dissociation rates within porous networks would further improve predictive capabilities. The application of deep learning architectures could also be explored to better capture nonlinearities inherent in complex pore structures.

With regard to hydrate transport in pipelines, the CFD–DEM simulations revealed critical insights into clustering, stratification, and pressure drop behavior. Future work should aim to simulate longer timescales to capture late-stage phenomena such as hydrate plug formation, channeling, or flow restoration events. Incorporating additional physical mechanisms such as hydrate particle agglomeration, fragmentation, and breakage would provide a more complete picture of slurry transport behavior. Furthermore, studying hydrate dynamics in more complex geometries such as bends, tees, risers, and variable diameter pipelines would yield practical insights relevant to real offshore systems. A dynamic coupling of hydrate formation and dissociation kinetics with CFD–DEM particle transport could also be developed to simulate more realistic transient flow scenarios.

Finally, although this thesis applied physics-based and machine learning approaches separately, there is growing interest in hybrid models that combine physical laws with data-driven learning. Future research could explore the development of physics-informed machine learning (PINNs) frameworks that enforce thermodynamic consistency within ML predictions. Machine learning models could also be used to accelerate thermodynamic simulations by providing fast surrogate models for expensive calculations. Interpretability methods such as SHAP could be applied across hybrid models to retain transparency and ensure that predictions align with physical understanding.

Pursuing these research directions will not only refine the current predictive models but also open new possibilities for safe, efficient, and innovative applications of gas hydrates in energy systems, carbon management, and environmental technologies. A multidisciplinary integration of experimental work, physical modeling, and advanced machine learning will be essential for future progress in this important field.

References

[1] Broseta D, Ruffine L, Desmedt A. Gas Hydrates 1: Fundamentals, Characterization and Modeling. John Wiley & Sons; 2017.

[2] Sloan Jr ED, Koh CA. Clathrate hydrates of natural gases. CRC press; 2007.

[3] Khokhar AA, Gudmundsson JS, Sloan ED. Gas storage in structure H hydrates. Fluid Phase Equilib 1998;150:383–92.

[4] Kinnari K, Hundseid J, Li X, Askvik KM. Hydrate management in practice. J Chem Eng Data 2015;60:437–46.

[5] Mimachi H, Takahashi M, Takeya S, Gotoh Y, Yoneyama A, Hyodo K, et al. Effect of long-term storage and thermal history on the gas content of natural gas hydrate pellets under ambient pressure. Energy & Fuels 2015;29:4827–34.

[6] Xu C-G, Li X-S. Research progress of hydrate-based CO 2 separation and capture from gas mixtures. Rsc Adv 2014;4:18301–16.

[7] Shahnazar S, Hasan N. Gas hydrate formation condition: Review on experimental and modeling approaches. Fluid Phase Equilib 2014;379:72–85.

[8] Lee Y, Kim H, Lee W, Kang DW, Lee JW, Ahn Y-H. Thermodynamic and kinetic properties of CO2 hydrates and their applications in CO2 capture and separation. J Environ Chem Eng 2023;11:110933.

[9] Misyura S, Strizhak P, Meleshkin A, Morozov V, Gaidukova O, Shlegel N, et al. A Review of Gas Capture and Liquid Separation Technologies by CO2 Gas Hydrate. Energies 2023;16:3318.

[10] Salehy Y, Clain P, Boufares A, Chami N, Torré JP, Dalmazzone D, et al. Rheological properties of CO2 hydrate slurries in presence of Dioctyl sodium sulfosuccinate (AOT) in a dynamic loop for refrigeration application. Chem Eng Sci 2023;275:118716.

[11] Yang K, Chen Z, Zhang P. State-of-the-art of cold energy storage, release and transport using CO2 double hydrate slurry. Appl Energy 2024;358:122531.

[12] Mok J, Choi W, Seo Y. Theoretically achievable efficiency of hydrate-based desalination and its significance for evaluating kinetic desalination performance of gaseous hydrate formers. Desalination 2022;524:115487.

[13] Montazeri SM, Kolliopoulos G. Hydrate based desalination for sustainable water treatment: A review. Desalination 2022;537:115855.

[14] Yin Z, Chong ZR, Tan HK, Linga P. Review of gas hydrate dissociation kinetic models for energy recovery. J Nat Gas Sci Eng 2016;35:1362–87.

[15] Vlasov VA. Formation and dissociation of gas hydrate in terms of chemical kinetics. React Kinet Mech Catal 2013;110:5–13.

[16] Abbasi GR, Arif M, Isah A, Ali M, Mahmoud M, Hoteit H, et al. Gas hydrate

characterization in sediments via x-ray microcomputed tomography. Earth-Science Rev 2022;234:104233.

[17] Hassanpouryouzband A, Joonaki E, Farahani MV, Takeya S, Ruppel C, Yang J, et al. Gas hydrates in sustainable chemistry. Chem Soc Rev 2020;49:5225–309.

[18] Bavoh CB, Lal B, Osei H, Sabil KM, Mukhtar H. A review on the role of amino acids in gas hydrate inhibition, CO2 capture and sequestration, and natural gas storage. J Nat Gas Sci Eng 2019;64:52–71.

[19] Javanmardi J, Moshfeghian M, Maddox RN. An accurate model for prediction of gas hydrate formation conditions in mixtures of aqueous electrolyte solutions and alcohol. Can J Chem Eng 2001;79:367–73.

[20] Nasrifar K, Moshfeghian M. A model for prediction of gas hydrate formation conditions in aqueous solutions containing electrolytes and/or alcohol. J Chem Thermodyn 2001;33:999–1014.

[21] Eslamimanesh A, Mohammadi AH, Richon D. Thermodynamic model for predicting phase equilibria of simple clathrate hydrates of refrigerants. Chem Eng Sci 2011;66:5439–45.

[22] Moradi G, Khosravani E. Application of PRSV2 equation of state to predict hydrate formation temperature in the presence of inhibitors. Fluid Phase Equilib 2012;333:18–26.

[23] Moradi G, Khosravani E. Modeling of hydrate formation conditions for CH4, C2H6, C3H8, N2, CO2 and their mixtures using the PRSV2 equation of state and obtaining the Kihara potential parameters for these components. Fluid Phase Equilib 2013;338:179–87.

[24] Van der Waals JH. The statistical mechanics of clathrate compounds. Trans Faraday Soc 1956;52:184–93.

[25] McKoy V, Sinanoğlu O. Theory of dissociation pressures of some gas hydrates. J Chem Phys 1963;38:2946–56.

[26] Eslamimanesh A, Gharagheizi F, Mohammadi AH, Richon D. Phase equilibrium modeling of structure H clathrate hydrates of methane+ water “insoluble” hydrocarbon promoter using QSPR molecular approach. J Chem Eng Data 2011;56:3775–93.

[27] Soroush E, Mesbah M, Shokrollahi A, Rozyn J, Lee M, Kashiwao T, et al. Evolving a robust modeling tool for prediction of natural gas hydrate formation conditions. J Unconv Oil Gas Resour 2015;12:45–55.

[28] Zhong H-Q, Yao Q-L, Wang Y, He Y-F, Li Z-H. A graphical alternating conditional expectation to predict hydrate phase equilibrium conditions for sweet and sour natural gases. Math Probl Eng 2019;2019.

[29] Xu H, Jiao Z, Zhang Z, Huffman M, Wang Q. Prediction of methane hydrate formation conditions in salt water using machine learning algorithms. Comput Chem Eng 2021;151:107358.

[30] Sinehbaghizadeh S, Javanmardi J, Roosta A, Mohammadi AH. Estimation of the dissociation conditions and storage capacities of various sH clathrate hydrate systems using

effective deterministic frameworks. Fuel 2019;247:272–86.

[31] Li H, Sun H, Chen J, Chen B, Zhong D, Yang M. Advances, Applications, and Perspectives of Machine Learning Approaches in Predicting Gas Hydrate Phase Equilibrium. Energy & Fuels 2024.

[32] Kang S-P, Lee J-W, Ryu H-J. Phase behavior of methane and carbon dioxide hydrates in meso-and macro-sized porous media. Fluid Phase Equilib 2008;274:68–72.

[33] Seo Y, Lee H, Uchida T. Methane and carbon dioxide hydrate phase behavior in small porous silica gels: three-phase equilibrium determination and thermodynamic modeling. Langmuir 2002;18:9164–70.

[34] Uchida T, Ebinuma T, Ishizaki T. Dissociation condition measurements of methane hydrate in confined small pores of porous glass. J Phys Chem B 1999;103:3659–62.

[35] Anderson R, Llamedo M, Tohidi B, Burgass RW. Experimental measurement of methane and carbon dioxide clathrate hydrate equilibria in mesoporous silica. J Phys Chem B 2003;107:3507–14.

[36] Zarifi M, Javanmardi J, Hashemi H, Eslamimanesh A, Mohammadi AH. Experimental study and thermodynamic modelling of methane and mixed C1+ C2+ C3 clathrate hydrates in the presence of mesoporous silica gel. Fluid Phase Equilib 2016;423:17–24.

[37] Uchida T, Ebinuma T, Takeya S, Nagao J, Narita H. Effects of pore sizes on dissociation temperatures and pressures of methane, carbon dioxide, and propane hydrates in porous media. J Phys Chem B 2002;106:820–6.

[38] Seo Y, Lee S, Cha I, Lee JD, Lee H. Phase equilibria and thermodynamic modeling of ethane and propane hydrates in porous silica gels. J Phys Chem B 2009;113:5487–92.

[39] Clarke MA, Pooladi-Darvish M, Bishnoi PR. A method to predict equilibrium conditions of gas hydrate formation in porous media. Ind Eng Chem Res 1999;38:2485–90.

[40] Duan X, Shi B, Wang J, Song S, Liu H, Li X, et al. Simulation of the hydrate blockage process in a water-dominated system via the CFD-DEM method. J Nat Gas Sci Eng 2021;96:104241.

[41] Sakurai S, Nakatsuka Y, Edwards TJ, Hoskin BJ, Manning DK. An experimental study for flow assurance of the methane hydrate production test system. Offshore Technol. Conf., OTC; 2014, p. D041S055R005.

[42] Song G, Li Y, Wang W, Jiang K, Shi Z, Yao S. Numerical simulation of pipeline hydrate particle agglomeration based on population balance theory. J Nat Gas Sci Eng 2018;51:251–61.

[43] Zhang L, Jin R, Wang J, Yao H, Li Y, Yin F, et al. Kinetic Properties and Dynamic Distribution of Hydrate Particles in a Water-Dominated System. Energy & Fuels 2024;38:18457–68.

[44] Wang Z, Li Z, Pei J, Ma N, Zhang J, Sun B. Simulation of Hydrate Particles Aggregation and Deposition in Gas-Dominated Flow. SPE J 2024;29:1492–509.

[45] Sloan ED. Clathrate hydrates: the other common solid water phase. Ind Eng Chem Res 2000;39:3123–9.

[46] Jeffrey GA. Hydrate inclusion compounds. J Incl Phenom 1984;1:211–22.

[47] Koh CA, Sum AK, Sloan ED. State of the art: Natural gas hydrates as a natural resource. J Nat Gas Sci Eng 2012;8:132–8.

[48] Chong ZR, Yang SHB, Babu P, Linga P, Li X-S. Review of natural gas hydrates as an energy resource: Prospects and challenges. Appl Energy 2016;162:1633–52.

[49] Veluswamy HP, Kumar A, Seo Y, Lee JD, Linga P. A review of solidified natural gas (SNG) technology for gas storage via clathrate hydrates. Appl Energy 2018;216:262–85.

[50] Qin Y, Shang L, Lv Z, He J, Yang X, Zhang Z. Methane hydrate formation in porous media: Overview and perspectives. J Energy Chem 2022;74:454–80. https://doi.org/https://doi.org/10.1016/j.jechem.2022.07.019.

[51] Bhattacharjee G, Veluswamy HP, Kumar A, Linga P. Stability analysis of methane hydrates for gas storage application. Chem Eng J 2021;415:128927.

[52] Omran A, Nesterenko N, Paecklar AA, Barrier N, Valtchev V. Toward Economical Seawater-Based Methane Hydrate Formation at Ambient Temperature: A Combined Experimental and Computational Study. ACS Sustain Chem Eng 2022;10:11617–26. https://doi.org/10.1021/acssuschemeng.2c03530.

[53] Eslamimanesh A, Mohammadi AH, Richon D, Naidoo P, Ramjugernath D. Application of gas hydrate formation in separation processes: A review of experimental studies. J Chem Thermodyn 2012;46:62–71.

[54] Zhang J, Wang Z, Liu S, Zhang W, Yu J, Sun B. Prediction of hydrate deposition in pipelines to improve gas transportation efficiency and safety. Appl Energy 2019;253:113521.

[55] Zhang J, Wang Z, Sun B, Sun X, Liao Y. An integrated prediction model of hydrate blockage formation in deep-water gas wells. Int J Heat Mass Transf 2019;140:187–202.

[56] Hammerschmidt EG. Formation of gas hydrates in natural gas transmission lines. Ind Eng Chem 1934;26:851–5.

[57] Motiee M. Estimate possibility of hydrates. Hydrocarb Process (International Ed) 1991;70:98–9.

[58] Jager MD, Sloan ED. The effect of pressure on methane hydration in pure water and sodium chloride solutions. Fluid Phase Equilib 2001;185:89–99.

[59] Towler BF, Mokhatab SBT-HP. Quickly estimate hydrate formation conditions in natural gases: using this simple technique to predict unfavorable conditions can save time and cost 2005;84:61+.

[60] Bahadori A, Vuthaluru HB. A novel correlation for estimation of hydrate forming condition of natural gases. J Nat Gas Chem 2009;18:453–7.

[61] Ghiasi MM. Initial estimation of hydrate formation temperature of sweet natural gases based on new empirical correlation. J Nat Gas Chem 2012;21:508–12.

[62] Mohamadi-Baghmolaei M, Hajizadeh A, Azin R, Izadpanah AA. Assessing thermodynamic models and introducing novel method for prediction of methane hydrate formation. J Pet Explor Prod Technol 2018;8:1401–12.

[63] Kummamuru NB, Perreault P, Lenaerts S. A new generalized empirical correlation for predicting methane hydrate equilibrium conditions in pure water. Ind Eng Chem Res 2021;60:3474–83.

[64] Medeiros F de A, Segtovich ISV, Tavares FW, Sum AK. Sixty years of the van der Waals and Platteeuw model for clathrate hydrates—A critical review from its statistical thermodynamic basis to its extensions and applications. Chem Rev 2020;120:13349–81.

[65] Khan MN, Warrier P, Peters CJ, Koh CA. Advancements in hydrate phase equilibria and modeling of gas hydrates systems. Fluid Phase Equilib 2018;463:48–61.

[66] Parrish WR, Prausnitz JM. Dissociation Pressures of Gas Hydrates Formed by Gas Mixtures. Ind Eng Chem Process Des Dev 1972;11:26–35. https://doi.org/10.1021/i260041a006.

[67] Chen G-J, Guo T-M. A new approach to gas hydrate modelling. Chem Eng J 1998;71:145–51.

[68] Chen X, Li H. New pragmatic strategies for optimizing Kihara potential parameters used in van der Waals-Platteeuw hydrate model. Chem Eng Sci 2022;248:117213.

[69] Mahabadian MA, Chapoy A, Burgass R, Tohidi B. Development of a multiphase flash in presence of hydrates: Experimental measurements and validation with the CPA equation of state. Fluid Phase Equilib 2016;414:117–32.

[70] Sirino TH, Neto MAM, Bertoldi D, Morales REM, Sum AK. Multiphase flash calculations for gas hydrates systems. Fluid Phase Equilib 2018;475:45–63.

[71] MohamadiBaghmolaei M, Mahmoudy M, Jafari D, MohamadiBaghmolaei R, Tabkhi F. Assessing and optimization of pipeline system performance using intelligent systems. J Nat Gas Sci Eng 2014;18:64–76.

[72] Garapati N, Anderson BJ. Statistical thermodynamics model and empirical correlations for predicting mixed hydrate phase equilibria. Fluid Phase Equilib 2014;373:20–8.

[73] Hosseini M, Rahimi R, Ghaedi M. Hydrogen sulfide solubility in different ionic liquids: an updated database and intelligent modeling. J Mol Liq 2020;317:113984.

[74] Hemmati-Sarapardeh A, Varamesh A, Husein MM, Karan K. On the evaluation of the viscosity of nanofluid systems: Modeling and data assessment. Renew Sustain Energy Rev 2018;81:313–29.

[75] Esmaeili S, Sarma H, Harding T, Maini B. A data-driven model for predicting the effect of temperature on oil-water relative permeability. Fuel 2019;236:264–77.

[76] Zarei MM, Hosseini M, Mohammadi AH, Moosavi A. Model development for estimating

calcium sulfate dihydrate, hemihydrate, and anhydrite solubilities in multicomponent acid and salt containing aqueous solutions over wide temperature ranges. J Mol Liq 2021;328:115473.

[77] Mesbah M, Soroush E, Rezakazemi M. Development of a least squares support vector machine model for prediction of natural gas hydrate formation temperature. Chinese J Chem Eng 2017;25:1238–48.

[78] Amar MN. Prediction of hydrate formation temperature using gene expression programming. J Nat Gas Sci Eng 2021;89:103879.

[79] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–30.

[80] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn 2006;63:3–42.

[81] De Roo JL, Peters CJ, Lichtenthaler RN, Diepen GAM. Occurrence of methane hydrate in saturated and unsaturated solutions of sodium chloride and water in dependence of temperature and pressure. AIChE J 1983;29:651–7.

[82] Dholabhai PD, Englezos P, Kalogerakis N, Bishnoi PR. Equilibrium conditions for methane hydrate formation in aqueous mixed electrolyte solutions. Can J Chem Eng 1991;69:800–5.

[83] Du J, Wang X, Liu H, Guo P, Wang Z, Fan S. Experiments and prediction of phase equilibrium conditions for methane hydrate formation in the NaCl, CaCl2, MgCl2 electrolyte solutions. Fluid Phase Equilib 2019;479:1–8.

[84] Haghighi H, Chapoy A, Tohidi B. Methane and water phase equilibria in the presence of single and mixed electrolyte solutions using the cubic-plus-association equation of state. Oil Gas Sci Technol l'IFP 2009;64:141–54.

[85] Hu Y, Lee BR, Sum AK. Phase equilibrium data of methane hydrates in mixed brine solutions. J Nat Gas Sci Eng 2017;46:750–5.

[86] Hu Y, Makogon TY, Karanjkar P, Lee K-H, Lee BR, Sum AK. Gas hydrates phase equilibrium with CaBr2 and CaBr2+ MEG at ultra-high pressures. J Nat Gas Eng 2017;2:42–9.

[87] Hu Y, Makogon TY, Karanjkar P, Lee K-H, Lee BR, Sum AK. Gas hydrates phase equilibria for structure I and II hydrates with chloride salts at high salt concentrations and up to 200 MPa. J Chem Thermodyn 2018;117:27–32.

[88] Kamari A, Hashemi H, Babaee S, Mohammadi AH, Ramjugernath D. Phase stability conditions of carbon dioxide and methane clathrate hydrates in the presence of KBr, CaBr2, MgCl2, HCOONa, and HCOOK aqueous solutions: Experimental measurements and thermodynamic modelling. J Chem Thermodyn 2017;115:307–17.

[89] Kang S-P, Chun M-K, Lee H. Phase equilibria of methane and carbon dioxide hydrates in the aqueous MgCl2 solutions. Fluid Phase Equilib 1998;147:229–38.

[90] Kharrat M, Dalmazzone D. Experimental determination of stability conditions of methane

hydrate in aqueous calcium chloride solutions using high pressure differential scanning calorimetry. J Chem Thermodyn 2003;35:1489–505.

[91] Hu Y, Lee K-H, Lee BR, Sum AK. Gas hydrate formation from high concentration KCl brines at ultra-high pressures. J Ind Eng Chem 2017;50:142–6.

[92] Li S, Wang J, Lv X, Ge K, Jiang Z, Li Y. Experimental measurement and thermodynamic modeling of methane hydrate phase equilibria in the presence of chloride salts. Chem Eng J 2020;395:125126.

[93] Maekawa T, Imai N. Equilibrium conditions of methane and ethane hydrates in aqueous electrolyte solutions. Ann N Y Acad Sci 2000;912:932–9.

[94] Maekawa T, Itoh S, Sakata S, Igari S, Imai N. Pressure and temperature conditions for methane hydrate dissociation in sodium chloride solutions. Geochem J 1995;29:325–9.

[95] Mohammadi AH, Afzal W, Richon D. Gas hydrates of methane, ethane, propane, and carbon dioxide in the presence of single NaCl, KCl, and CaCl2 aqueous solutions: Experimental measurements and predictions of dissociation conditions. J Chem Thermodyn 2008;40:1693–7.

[96] Mohammadi AH, Kraouti I, Richon D. Methane hydrate phase equilibrium in the presence of NaBr, KBr, CaBr2, K2CO3, and MgCl2 aqueous solutions: Experimental measurements and predictions of dissociation conditions. J Chem Thermodyn 2009;41:779–82.

[97] Hu Y, Makogon TY, Karanjkar P, Lee K-H, Lee BR, Sum AK. Gas hydrates phase equilibria and formation from high concentration NaCl brines up to 200 MPa. J Chem Eng Data 2017;62:1910–8.

[98] Aregbe AG, Sun B, Chen L. Methane Hydrate Dissociation Conditions in High-Concentration NaCl/KCl/CaCl2 Aqueous Solution: Experiment and Correlation. J Chem Eng Data 2019;64:2929–39.

[99] Adisasmito S, Frank III RJ, Sloan Jr ED. Hydrates of carbon dioxide and methane mixtures. J Chem Eng Data 1991;36:68–71.

[100] Atik Z, Windmeier C, Oellrich LR. Experimental gas hydrate dissociation pressures for pure methane in aqueous solutions of MgCl2 and CaCl2 and for a (methane+ ethane) gas mixture in an aqueous solution of (NaCl+ MgCl2). J Chem Eng Data 2006;51:1862–7.

[101] Bhawangirkar DR, Sangwai JS. Phase equilibrium of methane hydrates in the presence of MgBr2, CaBr2, and ZnBr2 aqueous solutions. J Chem Eng Data 2021;66:2519–30.

[102] Cha M, Hu Y, Sum AK. Methane hydrate phase equilibria for systems containing NaCl, KCl, and NH4Cl. Fluid Phase Equilib 2016;413:2–9.

[103] Chong ZR, Koh JW, Linga P. Effect of KCl and MgCl2 on the kinetics of methane hydrate formation and dissociation in sandy sediments. Energy 2017;137:518–29.

[104] Carroll J. Natural gas hydrates: a guide for engineers. Gulf Professional Publishing; 2020.

[105] Majid AAA, Worley J, Koh CA. Thermodynamic and kinetic promoters for gas hydrate technological applications. Energy & Fuels 2021;35:19288–301.

[106] Suginaka T, Sakamoto H, Iino K, Sakakibara Y, Ohmura R. Phase equilibrium for ionic semiclathrate hydrate formed with CO2, CH4, or N2 plus tetrabutylphosphonium bromide. Fluid Phase Equilib 2013;344:108–11.

[107] Muromachi S, Takeya S, Yamamoto Y, Ohmura R. Characterization of tetra-n-butylphosphonium bromide semiclathrate hydrate by crystal structure analysis. CrystEngComm 2014;16:2056–60.

[108] Liu Y, Wang X, Lang C, Zhao J, Lv X, Ge Y, et al. Experimental study on the gas hydrates blockage and evaluation of kinetic inhibitors using a fully visual rocking cell. J Nat Gas Sci Eng 2021;96:104331.

[109] Wang J, Meng Y, Han B, Liu Z, Zhang L, Yao H, et al. Hydrate blockage in subsea oil/gas flowlines: Prediction, prevention, and remediation. Chem Eng J 2023;461:142020.

[110] Hosseini M, Leonenko Y. Development of explicit models to predict methane hydrate equilibrium conditions in pure water and brine solutions: A machine learning approach. Chem Eng Sci 2024;286:119603.

[111] Mooijer-Van Den Heuvel MM, Peters CJ, de Swaan Arons J. Influence of water-insoluble organic components on the gas hydrate equilibrium conditions of methane. Fluid Phase Equilib 2000;172:73–91.

[112] Illbeigi M, Fazlali A, Mohammadi AH. Thermodynamic model for the prediction of equilibrium conditions of clathrate hydrates of methane+ water-soluble or-insoluble hydrate former. Ind Eng Chem Res 2011;50:9437–50.

[113] Kim S, Choi S-D, Seo Y. CO2 capture from flue gas using clathrate formation in the presence of thermodynamic promoters. Energy 2017;118:950–6.

[114] Hassan MHA, Sher F, Zarren G, Suleiman N, Tahir AA, Snape CE. Kinetic and thermodynamic evaluation of effective combined promoters for CO2 hydrate formation. J Nat Gas Sci Eng 2020;78:103313.

[115] Seo D, Lee S, Moon S, Lee Y, Park Y. Investigating two synthetic routes for gas hydrate formation to control the trapping of methane from natural gas. Chem Eng J 2023;467:143512.

[116] Jager MD, De Deugd RM, Peters CJ, de Swaan Arons J, Sloan ED. Experimental determination and modeling of structure II hydrates in mixtures of methane+ water+ 1, 4-dioxane. Fluid Phase Equilib 1999;165:209–23.

[117] De Deugd RM, Jager MD, de Swaan Arons J. Mixed hydrates of methane and water-soluble hydrocarbons modeling of empirical results. AIChE J 2001;47:693–704.

[118] Thakre N, Jana AK. Physical and molecular insights to Clathrate hydrate thermodynamics. Renew Sustain Energy Rev 2021;135:110150. https://doi.org/https://doi.org/10.1016/j.rser.2020.110150.

[119] Ismail I, Gaganis V, Marinakis D, Mousavi R, Tohidi B. Accuracy of different thermodynamic software packages in predicting hydrate dissociation conditions. Chem Thermodyn Therm Anal 2023;9:100103.

https://doi.org/https://doi.org/10.1016/j.ctta.2022.100103.

[120] Van der Waals JH. Clathrate solutions. Adv Chem Phys 1959;2:1–57.

[121] Ilani-Kashkouli P, Babaee S, Gharagheizi F, Hashemi H, Mohammadi AH, Ramjugernath D. An assessment test for phase equilibrium data of water soluble and insoluble clathrate hydrate formers. Fluid Phase Equilib 2013;360:68–76.

[122] Babaee S, Hashemi H, Javanmardi J, Eslamimanesh A, Mohammadi AH. Thermodynamic model for prediction of phase equilibria of clathrate hydrates of hydrogen with different alkanes, alkenes, alkynes, cycloalkanes or cycloalkene. Fluid Phase Equilib 2012;336:71–8.

[123] Mohammadi AH, Richon D. A mathematical model based on artificial neural network technique for estimating liquid water− hydrate equilibrium of water− hydrocarbon System. Ind Eng Chem Res 2008;47:4966–70.

[124] Lin Y, Zhou Z, Song Z, Shi Q, Hao Y, Fu Y, et al. Insights into the mechanical stability of tetrahydrofuran hydrates from experimental, machine learning, and molecular dynamics perspectives. Nanoscale 2024;16:6296–308.

[125] Zhang Y, Song Z, Lin Y, Shi Q, Hao Y, Fu Y, et al. Predicting mechanical properties of CO2 hydrates: machine learning insights from molecular dynamics simulations. J Phys Condens Matter 2023;36:15101.

[126] Lin Y, Hao Y, Shi Q, Xu Y, Song Z, Zhou Z, et al. Enhanced formation of methane hydrates via graphene oxide: Machine learning insights from molecular dynamics simulations. Energy 2024;289:130080.

[127] Shi Q, Lin Y, Hao Y, Song Z, Zhou Z, Fu Y, et al. Unconventional growth of methane hydrates: A molecular dynamics and machine learning study. Energy 2023;282:128337.

[128] Mohammadi AH, Richon D. Hydrate phase equilibria for hydrogen+water and hydrogen+tetrahydrofuran+water systems: Predictions of dissociation conditions using an artificial neural network algorithm. Chem Eng Sci 2010;65:3352–5. https://doi.org/https://doi.org/10.1016/j.ces.2010.02.015.

[129] Babakhani SM, Bahmani M, Shariati J, Badr K, Balouchi Y. Comparing the capability of artificial neural network (ANN) and CSMHYD program for predicting of hydrate formation pressure in binary mixtures. J Pet Sci Eng 2015;136:78–87.

[130] Mesbah M, Soroush E, Roham M, Shahsavari S. Phase equilibrium modeling of semi-clathrate hydrates of the CO2+ H2/CH4/N2+ TBAB aqueous solution system. Pet Sci Technol 2017;35:1588–94.

[131] Sinehbaghizadeh S, Roosta A, Rezaei N, Ghiasi MM, Javanmardi J, Zendehboudi S. Evaluation of phase equilibrium conditions of clathrate hydrates using connectionist modeling strategies. Fuel 2019;255:115649.

[132] Pornaroontham P, Kim K, Kulprathipanja S, Rangsunvigit P. Water-soluble organic former selection for methane hydrates by supervised machine learning. Energy Reports 2023;9:2935–46.

[133] Cheng C-T, Chen L-J, Lin S-T, Tang M, Chen P-C, Chen Y-P. Measurements for the dissociation conditions of methane hydrate in the presence of 2, 5-dihydrofuran and 3, 4-dihydro-2H-pyran. Fluid Phase Equilib 2013;338:114–8.

[134] Partoon B, Javanmardi J. Effect of Mixed Thermodynamic and Kinetic Hydrate Promoters on Methane Hydrate Phase Boundary and Formation Kinetics. J Chem Eng Data 2013;58:501–9. https://doi.org/10.1021/je301153t.

[135] Chapoy A, Burgass RW, Kalorazi BT. Hydrate inhibition in propylene glycol and glycerol systems. 8th Int. Conf. Gas Hydrates 2014, 2014.

[136] Lim D, Park S, Ro H, Shin K, Lee H. Effect of guest–host hydrogen bonding on thermodynamic stability of clathrate hydrates: Diazine isomers. J Phys Chem C 2015;119:10218–26.

[137] Juan Y-W, Tang M, Chen L-J, Lin S-T, Chen P-C, Chen Y-P. Measurements for the equilibrium conditions of methane hydrate in the presence of cyclopentanone or 4-hydroxy-4-methyl-2-pentanone additives. Fluid Phase Equilib 2015;386:162–7.

[138] Bavoh CB, Partoon B, Lal B, Keong LK. Methane hydrate-liquid-vapour-equilibrium phase condition measurements in the presence of natural amino acids. J Nat Gas Sci Eng 2017;37:425–34.

[139] Kumar A, Vedula SS, Kumar R, Linga P. Hydrate phase equilibrium data of mixed methane-tetrahydrofuran hydrates in saline water. J Chem Thermodyn 2018;117:2–8.

[140] Shu S-S, Tiwikrama AH, Yang C-D, Lee M-J. Phase equilibrium and dynamic behavior of methane hydrates decomposition via depressurization in the presence of a promoter tert-butanol. J Taiwan Inst Chem Eng 2019;95:119–30.

[141] Bhattacharjee G, Goh MN, Arumuganainar SEK, Zhang Y, Linga P. Ultra-rapid uptake and the highly stable storage of methane as combustible ice. Energy Environ Sci 2020;13:4946–61.

[142] Junthong S, Yodpetch V, Inkong K, Kulprathipanja S, Rangsunvigit P. Effects of pyrrolidine on methane hydrate formation: Thermodynamic, kinetic, and morphology perspectives. J Nat Gas Sci Eng 2021;96:104322.

[143] Mainusch S, Peters CJ, de Swaan Arons J, Javanmardi J, Moshfeghian M. Experimental determination and modeling of methane hydrates in mixtures of acetone and water. J Chem Eng Data 1997;42:948–50.

[144] Mohammadi AH, Richon D. Gas hydrate phase equilibrium in methane+ ethylene glycol, diethylene glycol, or triethylene glycol+ water system. J Chem Eng Data 2011;56:4544–8.

[145] Mohammadi AH, Richon D. Phase equilibria of methane hydrates in the presence of methanol and/or ethylene glycol aqueous solutions. Ind Eng Chem Res 2010;49:925–8.

[146] Mohammadi AH, Richon D. Phase equilibria of clathrate hydrates of tetrahydrofuran+ hydrogen sulfide and tetrahydrofuran+ methane. Ind Eng Chem Res 2009;48:7838–41.

[147] Maekawa T. Equilibrium conditions for clathrate hydrates formed from methane and

aqueous propanol solutions. Fluid Phase Equilib 2008;267:1–5.

[148] Anderson R, Chapoy A, Haghighi H, Tohidi B. Binary ethanol− methane clathrate hydrate formation in the system CH4-C2H5OH-H2O: phase equilibria and compositional analyses. J Phys Chem C 2009;113:12602–7.

[149] Seo Y-T, Kang S-P, Lee H. Experimental determination and thermodynamic modeling of methane and nitrogen hydrates in the presence of THF, propylene oxide, 1, 4-dioxane and acetone. Fluid Phase Equilib 2001;189:99–110.

[150] Shin W, Park S, Koh D-Y, Seol J, Ro H, Lee H. Water-soluble structure H clathrate hydrate formers. J Phys Chem C 2011;115:18885–9.

[151] Kuo P-C, Chen L-J, Lin S-T, Chen Y-P. Measurements for the dissociation conditions of methane hydrate in the presence of 2-methyl-2-propanol. J Chem Eng Data 2010;55:5036–9.

[152] Ro H, Shin W, Park S, Lee H. New gas hydrates promotors: Spectroscopic observation and phase equilibrium. ISOPE Int. Ocean Polar Eng. Conf., ISOPE; 2012, p. ISOPE-I.

[153] Lee Y-J, Kawamura T, Yamamoto Y, Yoon J-H. Phase Equilibrium Studies of Tetrahydrofuran (Thf)+ Ch4, Thf+ Co2, Ch4+ Co2, and Thf+ Co2+ Ch4 Hydrates. J Chem Eng Data 2012;57:3543–8.

[154] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 1988;28:31–6.

[155] Shrake A, Rupley JA. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. J Mol Biol 1973;79:351–71.

[156] RDKit: Open-source cheminformatics. https://www.rdkit.org n.d.

[157] Hosseini M, Leonenko Y. Prediction of hydrogen−brine interfacial tension at subsurface conditions: Implications for hydrogen geo-storage. Int J Hydrogen Energy 2024;58:485–94. https://doi.org/https://doi.org/10.1016/j.ijhydene.2024.01.227.

[158] Hosseini M, Leonenko Y. A reliable model to predict the methane-hydrate equilibrium: An updated database and machine learning approach. Renew Sustain Energy Rev 2023;173:113103.

[159] Yao Y, Qiu Y, Cui Y, Wei M, Bai B. Insights to surfactant huff-puff design in carbonate reservoirs based on machine learning modeling. Chem Eng J 2023;451:138022.

[160] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30.

[161] Zhang A-Y, Cai M, Wei N, Li H-T, Zhang C, Pei J, et al. Analysis of sensitivity to hydrate blockage risk in natural gas gathering pipeline. Pet Sci 2024.

[162] Ge B-B, Li X-Y, Zhong D-L, Lu Y-Y. Investigation of natural gas storage and transportation by gas hydrate formation in the presence of bio-surfactant sulfonated lignin. Energy 2022;244:122665.

[163] Zhang X, Liu Q, He J, Yuan Q, Li J, Wu Q, et al. Research progress of incremental synthesis and enhancement mechanism of natural gas hydrates: A review. Renew Sustain Energy Rev 2024;202:114695.

[164] Han G, Lee W, Kim M-K, Lee JW, Ahn Y-H. Hydrogen separation from hydrogen-compressed natural gas blends through successive hydrate formations. Chem Eng J 2024;483:149409.

[165] Petukhov AN, Atlaskin AA, Kudryavtseva MS, Shablykin DN, Stepanova EA, Zanozin ID, et al. Hydrate-based technique for natural gas processing: Experimental study of pressure-dropping and continuous modes. Chem Eng J Adv 2024;18:100596.

[166] Aminnaji M, Qureshi MF, Dashti H, Hase A, Mosalanejad A, Jahanbakhsh A, et al. CO2 Gas hydrate for carbon capture and storage applications–Part 1. Energy 2024:131579.

[167] Wang S, Sun H, Liu H, Xi D, Long J, Zhang L, et al. Novel vermiculite/tannic acid composite aerogels with outstanding CO2 storage via enhanced gas hydrate formation. Energy 2024;289:130033.

[168] Aldroubi SG, Zahid U, Abdelaziz OY, Fouad WA, Baaqeel H. Sustainable integration of a refrigerant hydrate-based desalination system with a dual-effect absorption refrigeration cycle. Sep Purif Technol 2025;361:131345.

[169] Sun L, Sun H, Wang T, Dong H, Zhang L, Yang L, et al. Self-driven and directional transport of water during hydrate formation: Potential application in seawater desalination and dewatering. Desalination 2023;548:116299.

[170] Lee W, Kang DW, Ahn Y-H, Lee JW. Blended hydrate seed and liquid promoter for the acceleration of hydrogen hydrate formation. Renew Sustain Energy Rev 2023;177:113217.

[171] Kim M-K, Ahn Y-H. Gas hydrates for hydrogen storage: a comprehensive review and future prospects. Korean J Chem Eng 2024;41:73–94.

[172] Sutter RM, Meunier V, Ruiz S, Doebelin J, Milo C, Rauh C, et al. Process-controlled gas hydrate formation in coffee solutions for the application of cold concentration. Innov Food Sci Emerg Technol 2024;96:103748.

[173] Pandey JS, Srivastava S, Feyissa AH, Tariq M, Tumba K. The potential role of gas hydrates: An emerging frontier in food science and engineering. J Food Eng 2024:112210.

[174] Wang F, Lv Y, Xia X, Wu X, Cheng C, Qi T, et al. Energy assessment and thermodynamic evolution of a novel semi-clathrate hydrate cold storage system with internally circulating gas bubble disturbance. Fuel 2023;353:129125.

[175] Xu W, Sun Z, Huang M. Enhancing hydrate formation for cold storage by isotridecanyl polyoxyethylene ether series surfactants in quiescent conditions. J Energy Storage 2024;90:111712.

[176] Ma K, Li D, Liang D. Reservoir stimulation technologies for natural gas hydrate: Research progress, challenges, and perspectives. Energy & Fuels 2023;37:10112–33.

[177] Lall D, Vishal V. Geophysical indicators and methods for producing oceanic gas hydrates.

Adv. Nat. Gas Form. Process. Appl. Vol. 3 Nat. Gas Hydrates, Elsevier; 2024, p. 307–23.

[178] Yusupov VI, Semiletov IP. Understanding the processes associated with the formation and decomposition of gas hydrates on the arctic shelf. Fuel 2025;388:134472.

[179] Parmentier F-JW, Thornton BF, Silyakova A, Christensen TR. Vulnerability of Arctic-Boreal methane emissions to climate change. Front Environ Sci 2024;12:1460155.

[180] You K, Flemings PB, Malinverno A, Collett TS, Darnell K. Mechanisms of methane hydrate formation in geological systems. Rev Geophys 2019;57:1146–96.

[181] Azimi A, Javanmardi J, Mohammadi AH. Development of thermodynamic frameworks for modeling of clathrate hydrates stability conditions in porous media. J Mol Liq 2021;329:115463.

[182] Mu L, Li X, Tan Q, Zhou Z, Zhao H, Cui Q. Thermodynamic modeling of gas hydrate phase equilibrium conditions in porous media in the presence of inhibitor solution. Fluid Phase Equilib 2023;568:113752.

[183] Sun R, Duan Z. An accurate model to predict the thermodynamic stability of methane hydrate and methane solubility in marine environments. Chem Geol 2007;244:248–62.

[184] Li X-S, Zhang Y, Li G, Chen Z-Y, Yan K-F, Li Q-P. Gas hydrate equilibrium dissociation conditions in porous media using two thermodynamic approaches. J Chem Thermodyn 2008;40:1464–74.

[185] Roodbari SK, Mohebbi V, Behbahani RM. Prediction of gas hydrates phase equilibrium in porous media–Pore size effect and thermodynamics approach. Fluid Phase Equilib 2025;593:114330.

[186] Mok J, Go W, Seo Y. Predicting Phase Equilibria of CO2 Hydrate in Complex Systems Containing Salts and Organic Inhibitors for CO2 Storage: A Machine Learning Approach. Energy & Fuels 2024;38:5322–33.

[187] Wu G, Coulon F, Feng J-C, Yang Z, Jiang Y, Zhang R. Machine learning models for fast selection of amino acids as green thermodynamic inhibitors for natural gas hydrate. J Mol Liq 2023;370:120952.

[188] Momeni K, Sadeghi MT, Fanaei MA. Prediction of methane and ethane hydrate phase equilibria in porous media using artificial neural network. Proc. Third Int. Conf. Model. Simul. Appl. Optim. Sharjah, UAE January 20À22, Citeseer; 2009.

[189] Mohammadi AH, Eslamimanesh A, Richon D, Gharagheizi F, Yazdizadeh M, Javanmardi J, et al. Gas hydrate phase equilibrium in porous media: mathematical modeling and correlation. Ind Eng Chem Res 2012;51:1062–72.

[190] Wan L, Cao P, Sheng J. Phase Stability of CH4 and CO2 Hydrates under Confinement Predicted by Machine Learning. Langmuir 2024;40:22058–67.

[191] Wilder JW, Seshadri K, Smith DH. Modeling hydrate formation in media with broad pore size distributions. Langmuir 2001;17:6729–35.

[192] Klauda JB, Sandler SI. Modeling gas hydrate phase equilibria in laboratory and natural

porous media. Ind Eng Chem Res 2001;40:4197–208.

[193] Chen L-T, Sun C-Y, Chen G-J, Nie Y-Q. Thermodynamics model of predicting gas hydrate in porous media based on reaction− adsorption two-step formation mechanism. Ind Eng Chem Res 2010;49:3936–43.

[194] Soave G. Equilibrium constants from a modified Redlich-Kwong equation of state. Chem Eng Sci 1972;27:1197–203.

[195] Klauda JB, Sandler SI. A fugacity model for gas hydrate phase equilibria. Ind Eng Chem Res 2000;39:3377–86.

[196] Hosseini M, Leonenko Y. Predicting interfacial tension in brine-hydrogen/cushion gas systems under subsurface conditions: Implications for hydrogen geo-storage. Int J Hydrogen Energy 2024;91:1394–406.

[197] Hosseini M, Boudreault R, Leonenko Y. Insights into predicting equilibrium conditions of clathrate hydrates of methane+ water-soluble hydrate former. Fuel 2025;384:133945.

[198] Tohidi B, Burgass RW, Danesh A, Østergaard KK, Todd AC. Improving the accuracy of gas hydrate dissociation point measurements. Ann N Y Acad Sci 2000;912:924–31.

[199] Kim D, Ahn Y-H, Kim S-J, Lee JY, Lee J, Seo Y, et al. Gas hydrate in crystalline-swelled clay: the effect of pore dimension on hydrate formation and phase equilibria. J Phys Chem C 2015;119:22148–53.

[200] Turner DJ, Cherry RS, Sloan ED. Sensitivity of methane hydrate phase equilibria to sediment pore size. Fluid Phase Equilib 2005;228:505–10.

[201] Smith DH, Wilder JW, Seshadri K. Methane hydrate equilibria in silica gels with broad pore-size distributions. AIChE J 2002;48:393–400.

[202] Klauda JB, Sandler SI. Predictions of gas hydrate phase equilibria and amounts in natural sediment porous media. Mar Pet Geol 2003;20:459–70.

[203] Sakurai S, Hoskin B, Choi J, May EF, Aman ZM, Norris BWE. Dynamics of methane hydrate particles in water-dominant systems during transient flow. Fuel 2022;324:124772.

[204] Liu Z, Sun B, Wang Z, Zhang J, Wang X. Prediction and management of hydrate reformation risk in pipelines during offshore gas hydrate development by depressurization. Fuel 2021;291:120116.

[205] Barker JW, Gomez RK. Formation of hydrates during deepwater drilling operations. J Pet Technol 1989;41:297–301.

[206] Wang J, Wang Q, Meng Y, Yao H, Zhang L, Jiang B, et al. Flow characteristic and blockage mechanism with hydrate formation in multiphase transmission pipelines: In-situ observation and machine learning predictions. Fuel 2022;330:125669.

[207] Liu X, Song K, Li Y, Yao H, Chen H, Li Z, et al. Evolution Process of Natural Gas Hydrate Deposits in Water-Dominated Systems. Energy & Fuels 2022;36:10003–15.

[208] Kong Q, Qin X, Ji P, Wang Z, Zhang J, Fu W, et al. Improved formation and evolution

models for hydrate flow barriers in water-dominated systems based on flowloop experiments. Chem Eng Sci 2024;297:120284.

[209] Yao M, Zhong Y, Qian Y, Xu Z, Xu T, Wang Z, et al. Growth and Evolution Characteristics and Stability of Hydrate Deposit Bed at Interfaces in Water-Dominated Systems. Energy & Fuels 2025.

[210] Sakurai S, Hoskin B, Choi J, Nonoue T, May EF, Kumar A, et al. Investigating hydrate formation rate and the viscosity of hydrate slurries in water-dominant flow: Flowloop experiments and modelling. Fuel 2021;292:120193.

[211] Sakurai S, Choi J, Manning N, Aman ZM, Nagaoka T, Omori M, et al. Rheology of methane-hydrate-slurry flow in 100% water cut system and influence of turbulent–laminar transition. Energy & Fuels 2024;38:8661–73.

[212] Wang Y, Koh CA, Dapena JA, Zerpa LE. A transient simulation model to predict hydrate formation rate in both oil-and water-dominated systems in pipelines. J Nat Gas Sci Eng 2018;58:126–34.

[213] Rensing PJ, Koh CA, Sloan ED. Simulation of hydrate aggregate structure via the discrete element method 2008.

[214] Liu J, Ning M, Dong T, Liang D. Numerical study on flow characteristics of hydrate slurry liquid-solid two-phase flow considering the adhesion between particles. J Nat Gas Sci Eng 2022;99:104410.

[215] Yao LM, Xiao ZM, Liu JB, Zhang Q, Wang M. An optimized CFD-DEM method for fluid-particle coupling dynamics analysis. Int J Mech Sci 2020;174:105503.

[216] Hellestø AS, Ghaffari M, Balakin B V, Hoffmann AC. A parametric study of cohesive particle agglomeration in a shear flow—Numerical simulations by the discrete element method. J Dispers Sci Technol 2017;38:611–20.

[217] Blais B, Lassaigne M, Goniva C, Fradette L, Bertrand F. Development of an unresolved CFD–DEM model for the flow of viscous suspensions and its application to solid–liquid mixing. J Comput Phys 2016;318:201–21.

[218] Wang J, Li Y, Zhao J, Shi B, Gong J, Li Q. Simulation of the effect of hydrate adhesion properties on flow safety in solid fluidization exploitation. Petroleum 2023;9:403–11.

[219] Sone Y, Sone Y. Kinetic theory and fluid dynamics. Springer; 2002.

[220] Blocken B, Stathopoulos T, Carmeliet J. CFD simulation of the atmospheric boundary layer: wall function problems. Atmos Environ 2007;41:238–52.

[221] Guo Y, Curtis JS. Discrete element method simulations for complex granular flows. Annu Rev Fluid Mech 2015;47:21–46.

[222] Kloss C, Goniva C, Hager A, Amberger S, Pirker S. Models, algorithms and validation for opensource DEM and CFD–DEM. Prog Comput Fluid Dyn an Int J 2012;12:140–52.

[223] Zhou ZY, Kuang SB, Chu KW, Yu A. Discrete particle simulation of particle–fluid flow: model formulations and their applicability. J Fluid Mech 2010;661:482–510.

[224] Nase ST, Vargas WL, Abatan AA, McCarthy JJ. Discrete characterization tools for cohesive granular material. Powder Technol 2001;116:214–23.

[225] Li L, Li B, Liu Z. Modeling of spout-fluidized beds and investigation of drag closures using OpenFOAM. Powder Technol 2017;305:364–76.

[226] Minglani D, Sharma A, Pandey H, Joshi JB. Analysis of flow behavior of cohesive monosized spherical and non-spherical particles in screw feeder. Powder Technol 2022;398:117049.