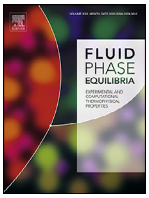
Fluid Phase Equilibria 530 (2021) 112894

Contents lists available at ScienceDirect

Fluid Phase Equilibria

journal homepage: www.elsevier.com/locate/fluid



Thermodynamic features-driven machine learning-based predictions of clathrate hydrate equilibria in the presence of electrolytes

Palash V. Acharya, Vaibhav Bahadur*

Walker Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX, United States

ARTICLE INFO

Article history:
Received 22 July 2020
Revised 27 October 2020
Accepted 4 November 2020
Available online 12 November 2020

Keywords:
Hydrate phase equilibrium
Decision tree
Thermodynamic activity
Electrolyte
Predictive modeling

ABSTRACT

Gas hydrates have significant applications in the areas of natural gas storage, desalination and gas separation. Knowledge of the thermodynamic conditions associated with hydrate formation is critical to their synthesis. Presently, we use machine learning (ML) to train and evaluate the performance of three algorithms on an experimental database (>1800 data points) to predict hydrate dissociation temperatures as a function of the constituent hydrate precursors and inhibitors. Importantly, and in contrast to most previous studies, we use thermodynamic variables such as the activity-based contribution due to electrolytes, partial pressure of individual gases, and specific gravity of the overall mixture as input features in the prediction algorithms. Using such features results in more physics-aware ML algorithms, which can capture the individual contributions of gases and electrolytes in a more fundamental manner. Three ML algorithms, Random Forest (RF), Extra Trees (ET), and Extreme Gradient Boosting (XGBoost) are employed and demonstrate excellent accuracy in their predictions of hydrate equilibrium conditions. The overall coefficient of determination (R^2) percentage is greater than 97% for all the ML models. XGBoost outperforms RF and ET with the highest overall coefficient of determination (R^2) and the lowest overall Average Absolute relative deviation (AARD) of 99.56% and 0.086% respectively.

1. Introduction

World energy consumption is forecast to increase by nearly 50% between 2018 and 2050 [1]. While the world continues its transition towards renewables, natural gas (methane) remains an attractive bridge fuel due to its relatively low carbon emissions upon combustion [2] when compared to other fossil fuels. Global natural gas consumption is projected to increase by more than 40% from 2018 to 2050 [1].

One of the biggest and largely untapped reservoirs of natural gas is in the form of hydrates, which are extensive in marine and permafrost environments. Gas hydrates are crystalline structures consisting of a cage of hydrogen-bonded water molecules, which trap a hydrocarbon molecule [3]. Gas hydrates form at high-pressure, low-temperature conditions. Hydrates represent a very attractive way of storing and transporting natural gas; 1 m^3 of solid hydrate can store up to 164 m^3 of methane (CH_4). It is estimated that just 15% of existing natural gas hydrate reserves can meet global energy demands for the next 200 years [4]. While commercial exploitation of hydrates will undoubtedly increase carbon emissions, hydrates are also being explored for carbon capture and sequestration (CCS), wherein atmospheric CO_2 is captured via synthesis of CO_2 hydrates [5–7]. Additional applications for hydrates are flow assurance, seawater desalination and gas separation [8,9].

Laboratory synthesis of gas hydrates is constrained by two factors: kinetics and thermodynamics. Kinetic constraints result in slow growth/conversion rates and very long induction/formation times. Thermodynamic constraints are about the thermodynamically stable pressure-temperature region required for hydrate formation. Hydrates-related applications are often governed by the combined interplay between both these factors. As an illustration, natural gas pipeline operators aim to prevent hydrate formation to avoid plugging. The addition of inhibiting chemicals such as electrolytes or alcohols shifts the thermodynamically stable region of hydrate formation to higher pressures and lower temperatures, which reduces the probability of hydrate formation. This study focusses on the thermodynamics aspect of hydrate formation. We predict the thermodynamically stable region of hydrate formation using machine learning techniques, which are grounded in fundamental thermodynamics, and which rely on an extensive experimental dataset.

Various methods proposed to predict hydrate dissociation temperatures (HDT) can be grouped into four categories. The first two are K-value method (uses vapour-solid equilibrium constants

* Corresponding author.
E-mail address: vb@austin.utexas.edu (V. Bahadur).

https://doi.org/10.1016/j.fluid.2020.112894

and K-value charts) [10] and gas gravity chart method (uses specific gravity of natural gases) [11] to predict hydrate formation/dissociation conditions. These approaches involve simple graphical techniques to estimate HDT and show significant deviations from experiments, especially for specific gravities between 0.9 and 1 [12]. Mixtures having the same specific gravity but different percentage compositions of constituent gases exhibit different equilibrium behaviour. This cannot be captured by the specific gravity method; errors of up to 50% have been reported [13]. The third method involves using empirical correlations containing parameters fitted to experimental data; these have been used mostly to predict hydrate forming conditions for sweet natural gases [14–18]. The fourth method involves statistical thermodynamics-based modeling approaches. The first such model developed in 1959, known as Van der Waals and Platteeuw (vdW-P) model was based on modification of classical adsorption statistical mechanics using Lennard-Jones potential function, to calculate dissociation pressure for pure gas hydrates [19]. Parrish and Prausnitz [20] extended this methodology for single and multicomponent gas hydrates by using this model along with the Kihara potential function to describe gas-water interactions. Several studies have used the vdW-P model coupled with various equations of state to calculate equilibrium conditions for gas hydrates in the presence of electrolytes and alcohols [21–29]. Such thermodynamic models used to characterize hydrate formation process require a very detailed knowledge of the underlying complex phenomena leading to hydrate formation. It is noted that there are significant challenges associated with such thermodynamic models that try to capture the physics underlying hydrate formation using macroscopic constructs, whereas hydrate formation events occur at a molecular level [30].

With the advent of computing technology, a fifth method, based on machine learning approaches, is rapidly gaining traction. Machine learning (ML) is increasingly being used to predict the behaviour of complex non-linear systems in fields ranging from finance, medicine, geology, sensors etc. [31–33]. ML algorithms, which are a subset of artificial intelligence, can predict performance and discern patterns characterizing a system by learning from data. They can also be used to model complex systems and automate analytical model building. Multiple machine learning algorithms such as neural networks, decision trees, and support vector machines have been recently studied to predict dissociation/formation conditions of hydrates. Such ML models are computationally fast and easier to implement when compared to conventional thermodynamic models.

The first study (in 1998) on this topic employed neural networks to predict hydrate dissociation conditions [34]. Four models employing different input features such as gas specific gravity and composition of hydrocarbons and inhibitors were studied using Artificial Neural Networks (ANN) [34]. Since then, multiple neural network based-studies employing specific gravity of the gas and either the pressure or temperature as the input variable have been used to predict hydrate forming conditions [35,36]. It is noted that the use of specific gravity as a variable results in loss of information on the gas composition. Gas composition has therefore been used as the input feature in a majority of ML-based predictions [37–45]. Recently, Hamidreza and Mohammad employed Extremely randomized trees and Least square support vector machines (LSSVM) on a database of more than 1840 experimental data points and achieved good predictions with an R^2 accuracy greater than 96% [40]. LSSVM and Gradient Boosted Regression trees have been employed in other studies to predict hydrate dissociation conditions with good accuracy [39,41,42].

Next, we highlight key limitations in existing studies that use ML techniques to predict hydrate dissociation/formation. Firstly, we note that a majority of prior studies employing ML to predict hydrate dissociation/formation conditions use total pressure and percentage composition of individual gases and additives as features to train the models. Exceptions include a few studies which have used either specific gravity [46] or electrical conductivity [47] of the aqueous solution as features. While models using percentage composition of gases and total pressure as features report good prediction capability, this approach makes it challenging to quantify the contribution of individual gases (in a mixture) towards hydrate formation. As an illustration, the same total pressure can be achieved for a 30% CO_2: 70% CH_4 mixture and a 70% CO_2: 30% CH_4 mixture (composition in molar volumes). In such cases, the total pressure feature in itself is of limited use in identifying the relative contribution of gases. The model would then primarily rely on the percentage composition of individual gases to predict the HDT. Secondly, in most existing studies, the contribution of inhibitors such as salts is captured via their individual weight percentage or molality term. From a chemistry standpoint, the inhibition action of salts depends on more fundamental factors such as the ionic strength of the solution, salt-water interactions etc. which cannot be captured via a simple weight percentage term. Thirdly, since ML models are data driven, the capability to predict HDT is highly contingent on the availability of experimental data for a particular salt or a combination of salts. As an illustration, the inhibiting influence of NaCl has been more widely studied than that of KCl, $CaCl_2$ or $MgCl_2$. This makes the database highly skewed towards prediction of NaCl-based inhibition.

This study includes multiple advancements in the use of ML algorithms for predictions of hydrate formation/dissociation conditions. Firstly, we individualize the contribution of every gas by calculating partial pressures and use them as features in ML models. While this may not necessarily lead to an improvement in model accuracy when compared to using molar composition and total pressure as features, it enables the algorithm to map hydrate dissociation temperatures to the fundamental parameters directly affecting the physics of the problem. This enables the algorithm to be paired up with models such as SHAP [48,49] that can track the relative contribution of every gas and salt towards hydrate dissociation temperature. This can provide valuable insights for designing processes in applications involving gas hydrates. Secondly, we calculate the contribution of the activity of water due to the presence of various electrolytes and use that as a feature in the ML model. The inhibiting influence of electrolytes on hydrate dissociation temperatures in thermodynamic models is captured via the activity of water which is defined as the ratio of the vapour pressure of a water sample and the vapour pressure of pure water at the same temperature [50,51]. Using activity as a feature allows us to capture the influence of both molality and the ionic strength while taking into consideration various intermolecular interactions. Another advantage of this approach stems from the fact that the weight percentages of different salts are fused into a single activity term. Lack or discontinuity in experimental data for any particular salt would be compensated by the activity term of another salt, thereby making the databank more efficient and continuous. Furthermore, activity being a more fundamental parameter, it could then be ascribed to reflect the presence of any salt or a combination of salts depending on the weight percentage leading to the same activity. In essence, this allows meaningful predictions of HDT from the use of a salt or a combination of salts. Thirdly, we use specific gravity as a feature to capture the influence of average molecular weight on the HDT. Overall, *we develop ML models using fundamental thermodynamic parameters of the constituent gases or chemicals in the system*. We use a low number of features, to be able to backtrack individual contributions of gases or salts; this would not be possible from the previously used approaches involving total pressures and weight percentages.

Presently, we evaluate the prediction performance of three different ensemble-based ML methods: Random Forest (RF) [52], Ex-

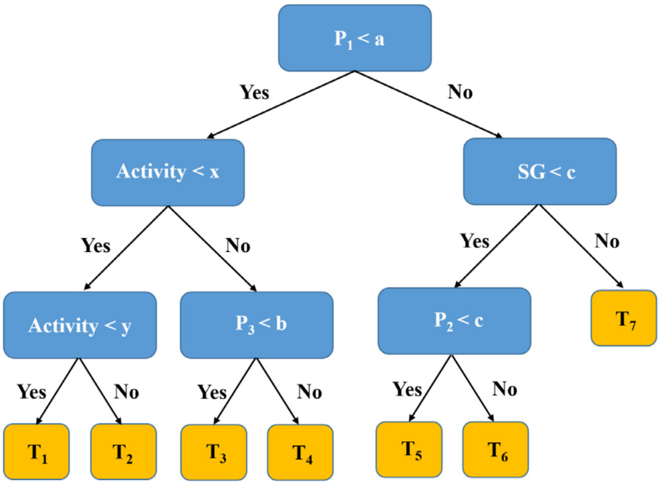


Fig. 1. An illustrative decision tree associated with hydrate dissociation temperature predictions. For any data point, these would be determined by the partial pressure of the individual gas components, specific gravity (SG) and the activity contribution due to the electrolyte, based on the splitting rules defined by a decision tree.

tremely randomized trees (ET) [53], and Extreme Gradient Boosting (XGBoost) [54]. We compare their relative performance to predict HDT using activity, partial pressure and specific gravity as input features. While prior studies have used gradient boosting (GB) algorithms, this is the first reported use of Extreme Gradient Boosting [54] (which is a computationally efficient variant of gradient boosting algorithm) for hydrates-related predictions. It is noted that due to its efficiency and ease of use, XGBoost (since its inception in 2015) has been widely used in Kaggle competitions and a variety of ML and data mining challenges.

2. Description of machine learning models used in this study

Ensemble-based techniques combine several base models to produce one optimal prediction model by decreasing variance and bias via bagging (bootstrap aggregating) and boosting. Amongst ensemble models, decision tree-based methods have become widely popular owing to their ease of implementation, versatility, and intuitive interpretation.

2.1. Random Forest and Extra Trees

A typical decision tree stratifies or segments a predictor space into several simple regions based on a set of splitting rules to optimize a specific objective function. The value of a prediction for a particular observation would be the mean or mode of the trained observations in the region/branch where it belongs (usually denoted as the terminal node or leaves of the tree) based on the splitting rules defining that particular observation. Since the set of splitting rules used to divide the predictor space can be displayed graphically as a tree (Fig. 1), such methods are referred to as decision tree-based methods.

Random forest utilizes two powerful ML techniques-bagging [55] and random feature selection [56] to provide a robust decision tree-based ensemble model. The ideology underlying random forest is as follows: fitting a single decision tree on a data set would cause the model to overfit the underlying trend, leading to high variance, wherein the model yields excellent predictions on the trained data set but performs poorly on unseen data. Bagging or bootstrap aggregating is therefore conducted on a data set to reduce variance and produce a model that can provide generalized predictions on any data set. In bagging, N different trees are fitted over N bootstrapped samples of data, and the results are averaged over all trees to obtain the final output. Random forest adds an extra layer of improvement over bagged regression trees by decorrelating individual trees in the forest. This is carried out by considering only a random subset of features every time a split decision is executed in an internal node. Extremely randomized trees are similar to random forest based methods, albeit with two key differences, i) the entire dataset is used to grow a tree instead of bootstrap sampling, ii) the internal cut point used to make a decision split is selected at random instead of searching for the most optimal split.

2.2. XGBoost (eXtreme Gradient Boosting)

Boosting is a form of additive modeling which is based on building a sequence of multiple weak learning models and combining them into a single composite strong model. The underlying ideology here is that the resultant collective model becomes a stronger predictor as more weak learning models are sequentially added to it. While the weak learning models can be built independently as in the case of random forest or extra trees, boosting relies on building these weak learners in a stage-wise fashion with each learner chosen to improve the overall model performance by optimizing a specific objective function.

In this study, we implement XGBoost algorithm [54] (operating under the framework of Gradient boosting), which uses the first and second derivative of the loss function to converge to global optimality quicker, while also improving the efficiency of the optimal solution of the model. The objective function minimized by XGBoost is as follows [54]:

$$obj(\theta) = \sum_i l\left(\hat{y}_i, y_i\right) + \sum_k \Omega(f_k) \tag{1}$$

$$\Omega(f_k) = \gamma' T + \frac{1}{2}\lambda' \|w\|^2 \tag{2}$$

where, l is a differentiable convex loss function that measures the difference between the prediction $\hat{y}_i$ and target y_i, and f_k is the k^{th} tree. The second term Ω helps to prevent overfitting by penalizing the complexity of the model in terms of the number of leaves in the tree T and vector of scores on leaves w. Here λ' is a regularized parameter and γ' is the learning rate, whose values lies between 0 and 1.

Since a tree ensemble model includes functions as parameters, it cannot be optimized using traditional optimization methods in Euclidean space and is therefore trained in an additive manner. The objective function to be minimized is then given by:

$$L^{(t)} = \sum_{i=1}^{k} \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + f_t(x_i) \right] + \Omega(f_t) \tag{3}$$

where $\hat{y}_i^t$ is the prediction of the i^{th} instance at the t^{th} iteration, and k is the total number of predictions. Therefore, the loss function is represented as the sum of the loss functions for the prediction till the *t-1*th iteration and a tree structure that, when added at the t^{th} iteration, most improves the model as per Eq. (3). Accordingly, the objective function can be optimized by using the second-order Taylor's approximation of the loss function (instead of first-order in general gradient boosting) which is given by [54]:

$$L^{(t)} \simeq \sum_{i=1}^{k} \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i {f_t}^2(x_i) \right] + \Omega(f_t) \tag{4}$$

where, $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = \partial^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ represent the first and second derivatives of each sample. The sum of loss values determines the loss function in Eq. (4) for each data sample corresponding to every leaf node. Assuming that the loss function is the mean square error function for regression problems and removing the constants, the objective function can be written for

Table 1
Summary of the experimental database employed in the present study.

Variable	Units	Min	Max	Average
P	MPa	0.13	72.26	6.92
T	K	247.52	303.48	278.35
CO_2	mol%	0	100	20.1
CH_4	mol%	0	100	51.59
C_2H_6	mol%	0	100	9.39
C_3H_8	mol%	0	100	12.67
n-C_4H_{10}	mol%	0	100	1.29
i-C_4H_{10}	mol%	0	88.8	1.26
N_2	mol%	0	100	3.69
NaCl	wt%	0	24.11	2.48
KCl	wt%	0	20	0.69
$CaCl_2$	wt%	0	25.75	1.14
$MgCl_2$	wt%	0	15	0.41

regression tree-based problems as [54]:

$$L^{(t)} \simeq \sum_{j=1}^{T}\left[\left(\sum_{i\in I_j} g_i\right)w_j + \frac{1}{2}\left(\sum_{i\in I_j} h_i + \lambda'\right){w_j}^2\right] + \gamma T \tag{5}$$

where, I_j represents all the data samples in the leaf node. During the process of building a tree, a particular node split will only be carried out if there is an improvement in the performance of the model as evaluated by this objective function.

3. Development of machine learning models

3.1. Experimental database

The predictive utility of any data-driven ML model is highly dependent on the experimental dataset employed for training the model. In order to predict the HDT of gas hydrates from pure gases and mixture of gases in the presence of electrolytes, an extensive dataset comprising of more than 1800 phase equilibrium data points was collected from published literature. Table 1 summarizes the key details of the gathered experimental dataset. The dataset includes P-T data for varied combinations of seven gases and four salts.

The data used in the present study was obtained from the following sources: Dholobhai et al. [57], Englezos and Ngan [58], Dholabhai and Bishnoi [59], Mei et al. [60], Kang et al. [61], Tatsuo [62], Jager and Sloan [63], Kharrat and Dalmazzone [64], Atik et al. [65], Mohammadi et al. [66], Haghighi et al. [67], De Roo et al. [68], Maekawa et al. [69], Dholabhai et al. [70], Nakane et al. [71], Tohidi et al. [72], Holder and Grigoriou [73], Paranjpe et al. [74], Adisasmito et al. [75], Adisasmito and Sloan [76], Nixdorf and Oellrich [77], Ng and Robinson [78], Holder and Hand [79], Ghavipour et al. [38], Jhaveri and Robinson [80], Wu et al. [81], Deaton and Frost [82], Verma [83], Unruh and Katz [84], Ohgaki et al. [85], Fan and Guo [86], Seo et al. [87] and Ng et al. [88].

3.2. Calculations of activity and partial pressures

In the present study, the contribution of electrolytes towards the activity of water is calculated using the Pitzer-Debye Huckel equation [89] and N-NRTL-NRF model [90] to capture the long and short range interactions, respectively. Accordingly, the contribution due to electrolytes can be evaluated via an activity term as:

$$\ln a_{w,el} = \sum_{i}^{N_{el}} \upsilon_i m_i \ln a_{w,el\,i} \Big/ \sum_{i}^{N_{el}} \upsilon_i m_i \tag{6}$$

where υ is the stoichiometric number of ions in the i^{th} electrolyte and m_i is the molality of the i^{th} electrolyte in the mixed electrolyte solution.

$$a_{w,eli} = x_w \gamma_w \tag{7}$$

where the activity coefficient of water γ_w is calculated using a combination of short range (SR) and long range (LR) interaction terms:

$$\ln\gamma_w = \ln\gamma^{SR} + \ln\gamma^{LR} \tag{8}$$

The short range contribution can be calculated using N-NRTL-NRF model as [90]:

$$\ln\gamma^{SR} = {x_{el}}^2\left(\lambda_{el,w}{\Gamma_{el,w}}^2 + \frac{\lambda_{el,w}{\Gamma_{el,w}}^2}{\beta_{w,el}} - \lambda_{el,w} - \lambda_{w,el}\right) \tag{9}$$

$$m_{eli} = \frac{1}{\upsilon_i}\sum_{j}^{N_{el}} \upsilon_j m_j \tag{10}$$

$$x_w = \frac{1000/M_w}{1000/M_w + \upsilon m_{eli}} \tag{11}$$

$$x_{el} = 1 - x_w \tag{12}$$

$$\upsilon = \upsilon_a + \upsilon_c \tag{13}$$

$$\Gamma_{w,el} = \frac{x_w \beta_{w,el}}{x_w \beta_{w,el} + x_{el}} \tag{14}$$

$$\Gamma_{el,w} = \frac{x_{el}\beta_{el,w}}{x_{el}\beta_{el,w} + x_w} \tag{15}$$

$$\beta_{w,el} = \exp\left(-\alpha\lambda_{w,el}\right) \tag{16}$$

$$\beta_{el,w} = \exp\left(-\alpha\lambda_{el,w}\right) \tag{17}$$

In the above equations, x_w & M_w denote the mole fraction and molecular weight of water, and m_{eli} is the molality of the i^{th} electrolyte in the single electrolyte solution. λ represents the optimized binary parameters for salt-water interactions (values for which have been taken from [91]), Γ denotes non-random factors and β is the Boltzmann factor. Note that interactions between dissolved gas and electrolytes has not been considered while evaluating $a_{w,eli}$. For the LR contribution, the Pitzer Debye Huckel equation for the ionic species can be expressed as [89]:

$$\ln\gamma^{LR} = \frac{x_{el}}{x_w}A_\phi\left(\frac{1000}{M_w}\right)^{1/2}\left(\frac{|z_a z_c| I^{1/2} - 2I^{3/2}}{I + \rho I^{1/2}}\right) \tag{18}$$

$$I = \frac{1}{2}\left(x_a {z_a}^2 + x_c {z_c}^2\right) \tag{19}$$

where I is the molar fraction ionic strength. The non-randomness factor α, Debye–Hückel constant A_ϕ, and closest approach ρ are equal to 0.125, 0.390947 and 14.90 respectively [90,92].

It is noted that we do not presently consider the influence of alcohol/s or other organic solvents on hydrate formation. We noticed that the activity calculated for alcohol/s using the commonly employed Margules equation [93] does not lead to the same extent of suppression in hydrate formation conditions when compared to electrolyte solutions with the same activity (calculated using Pitzer-Debye Huckel equation and N-NRTL-NRF model). This suggests that although the use of activity as a feature is an improvement over previous approaches, the models used to evaluate activity are also crucial to accurate prediction of hydrate forming conditions. To ensure consistency and obtain meaningful predictions from the present study, we considered only electrolytes in the analysis. It is noted that partial pressures of individual gases were estimated using mole fractions and total pressure assuming ideal gas behaviour.

Overall, the use of activity, specific gravity, and partial pressures as features instead of weight percentages and mole fractions makes the present approach more firmly grounded in fundamental chemical thermodynamics, and increases the generality and applicability of the predictive models.

Table 2
Accuracy of predictions obtained by ML algorithms employed in the present study.

	Random Forest			Extra Trees			XGBoost		
	Train	Test	Total	Train	Test	Total	Train	Test	Total
R^2%	97.84	94.35	97.53	98.45	94.08	98.01	99.85	98.06	99.56
AARD%	0.210	0.363	0.233	0.184	0.383	0.214	0.064	0.212	0.086

3.2. Procedure for conducting ML analysis

The dataset was divided into training and test data using train_test_split function from Scikit-learn library [94] which randomly assigns 85% of the dataset for training the model and the remaining 15% for evaluating the accuracy and performance of the trained model. The performance of a ML model depends significantly on its hyperparameters, which define the model's complexity and thereby its capacity to learn from any data. In tree-based models, typical hyperparameters include number of trees to grow in the forest, maximum depth of each tree, number of samples to be considered for each split, minimum number of data points to be allowed in a leaf node etc. For XGBoost algorithm, additional hyperparameters come into play, such as the learning rate, minimum loss reduction required to make a split, subsample ratio of the training instances etc. It is noted that tuning the hyperparameters using the training dataset without resorting to cross-validation can lead to overfitting where the model performs really well on the training data set but rather poorly on the test data set.

Presently, we resort to exhaustive grid search cross-validation (CV) with a two-step approach to carry out optimization. In the first step, a hyperparameter grid consisting of a wide range of parameter values was created, and combinations were sampled at random to narrow down a range of values, by evaluating their results using 5-fold cross-validation. Following this, an extensive grid search was conducted on the concentrated parameter space by evaluating all possible combinations in the hyperparameter grid to arrive at the combination yielding the best results. It is noted that the hyperparameters were evaluated to avoid overfitting the data set so as to provide the best fit for a broad spectrum of data points. To carry out grid search cross validation and implement the ML algorithms in Python, an open-source ML library Scikit-learn was used [94].

4. Results and discussions

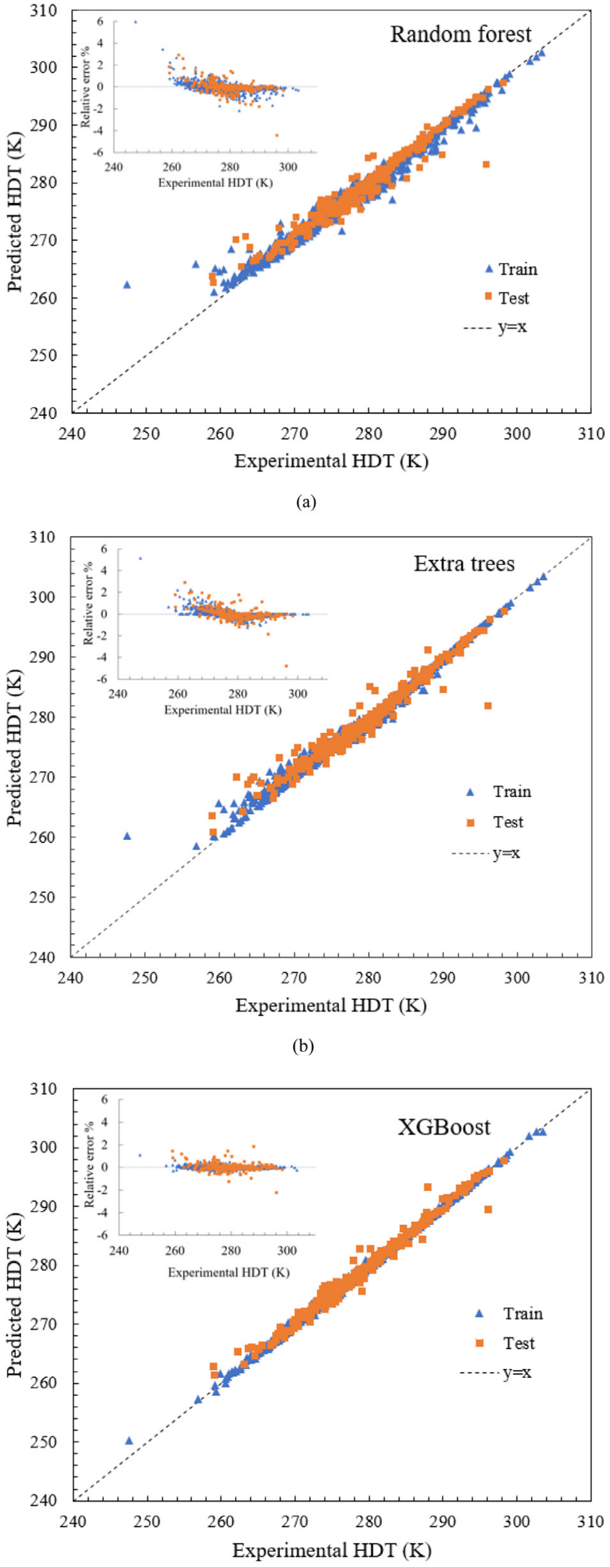
4.1. Validity of the model

After estimating optimum hyperparameters, ML algorithms were trained using thermodynamic variables (partial pressure, specific gravity and activity) as features. Subsequently, their prediction performance was evaluated on the test data set. In order to evaluate the reliability and accuracy of the developed ensemble models, two key statistical metrics were used, namely coefficient of determination percent (R^2%) and average absolute relative deviation percent (AARD%), as defined below:

$$R^2\% = 100\left(1 - \frac{\sum_{i=1}^{n}\left(T_{pred} - T_{\exp}\right)^2}{\sum_{i=1}^{n}\left(T_{pred} - average\left(T_{\exp}\right)\right)^2}\right) \tag{20}$$

$$AARD\% = 100\sum_{i=1}^{n}\left|\frac{T_{pred} - T_{\exp}}{T_{\exp}}\right|/n \tag{21}$$

Table 2 summarizes the performance metrics for Random forest, Extra Trees, and XGBoost models with the previously described experimental dataset of more than 1800 data points. It is evident

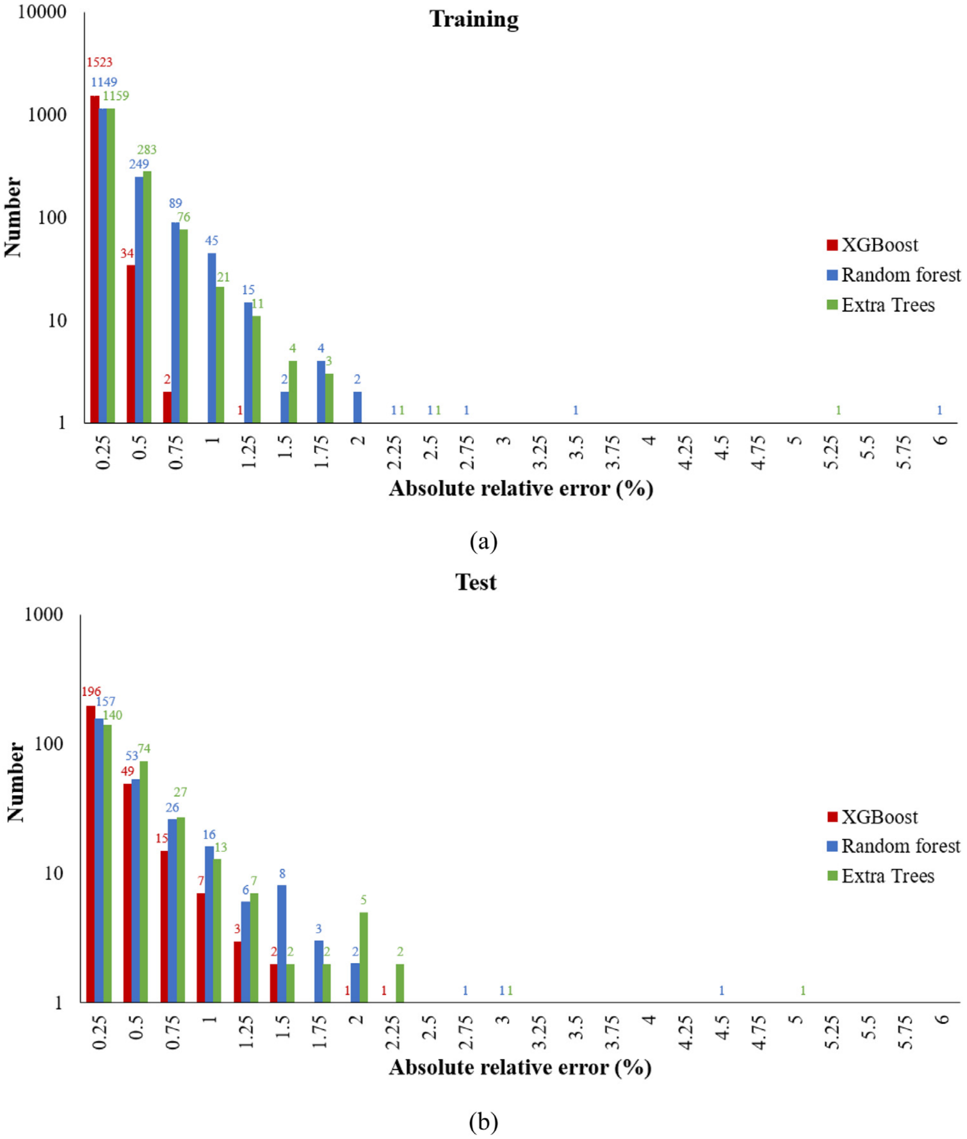


(a)

(b)

(c)

Fig. 2. Comparison of experimental hydrate dissociation temperatures versus predictions obtained from three ML models, (a) Random Forest, (b) Extra Trees, and (c) XGBoost.



(a)

(b)

Fig. 3. Distribution of the absolute relative error in predicting hydrate dissociation temperature (logarithmic y-axis) for (a) training and (b) test data sets obtained using XGBoost, Random Forest and Extra Trees algorithms.

that the XGBoost algorithm outperforms Random Forest and Extra Trees algorithms with R^2 accuracies of 99.85% (training data), 98.06% (test data) and 99.56% (overall). Random Forest and Extra Trees models exhibited almost similar prediction accuracies. The XGBoost algorithm results in a noticeably improved prediction accuracy for the test data (an improvement of about 4%) when compared to the other two algorithms. These results are in line with previous observations wherein boosting methods have exhibited improvements in predictions when compared to traditional models [95]. Importantly, all three ML models provide excellent prediction accuracies while using fundamental thermodynamic variables of the constituent gases and salts (activity, partial pressures, and specific gravity). This shows that setting up ML models in terms of fundamental thermodynamic parameters, as opposed to specifying compositions does not compromise the accuracy of the predictions.

Fig. 2 compares the predicted HDT for training and test data sets (combined) versus experimental data, the dashed line being the identity line, $y = x$. The distance of a point from the identity line is a measure of the prediction accuracy of the particular model. A distribution of points closely clustered to the identity line is a visual indicator of good predictions. It is clearly seen from Fig. 2c that XGBoost results in better predictions, as evident by a majority of data points very closely clustered around the identity line. The relative error is highest for two data points having the highest deviation from the identity line. These correspond to T_{exp} of 247.52 K [for $P_{methane}$ = 2.56 MPa (97.07 mol%) and $P_{propane}$ = 0.077 MPa (2.93 mol%)] and T_{exp} of 296.07 K [for $P_{carbon\ dioxide}$ = 1.91 MPa (8.09 mol%), $P_{methane}$ = 21.07 MPa (89.4 mol%) and P_{ethane} = 0.587 MPa (2.49 mol%)]. Deviations for these points can be very clearly noticed in the plots for Random Forest (Fig. 2a) and Extra Trees (Fig. 2b) methods. The deviation for the point T_{exp} = 247.52 K belonging to training data set can be attributed to a lack of experimental data around lower temperature ranges, which results in poorer predictions compared to other points. The deviation for the point T_{exp} = 296.07 K belonging to test data set at very high partial pressures of CH_4 can be explained either by a lack of experimental data points at higher pressures or possible local overfitting due to random allocation of test train data. It is noted that XGBoost results in better predictions at these two most deviant points as well. The relative deviation resulting from the use of XGBoost (1.12% and −2.23%, respectively) is much lower than the relative deviations resulting from the pre-

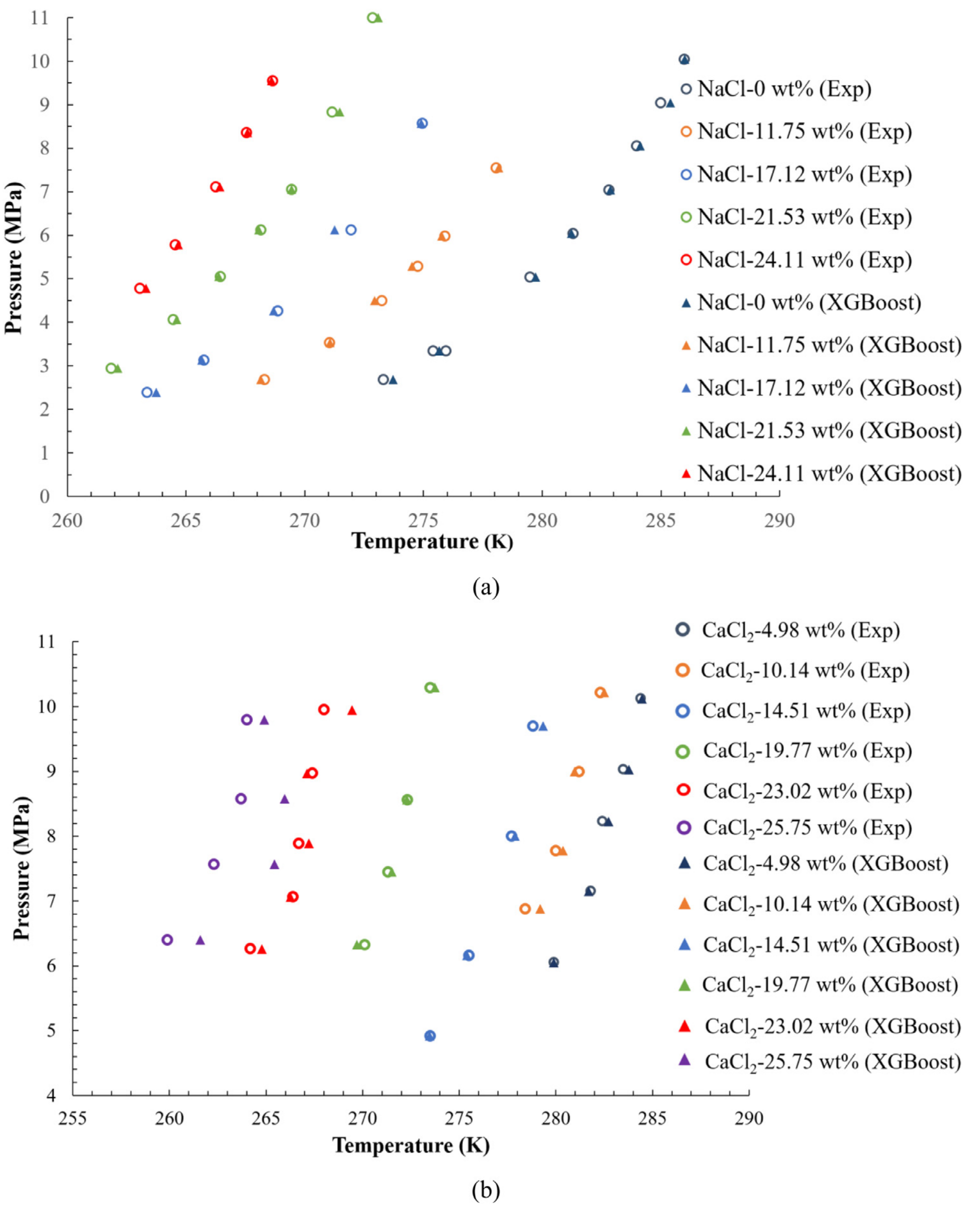


Fig. 4. Pressure-Temperature data associated with formation of methane hydrates in the presence of (a) NaCl and (b) $CaCl_2$ (varying weight percentages). Solid symbols represent the predictions from XGBoost model, while non-solid symbols represent experimental data [64,68].

diction of the Random Forest (5.98% and −4.40%, respectively) and Extra Trees (5.17% and −4.79%, respectively) methods. This highlights the robustness and superior predictive capabilities of the XGBoost model. It is noted that two data points corresponding to HDT of 242.09 K and 252.8 K, which belong to the lower end of the temperature spectrum, yielded absolute relative errors greater than 10% and have been considered as outliers in the present study.

Further insights on the relative performance of the models is obtained by analysing the distribution of the absolute relative error between the experimental and predicted HDTs for training and test data sets, as shown in Fig. 3 and the distribution of relative error, as shown in the inset of Fig. 2. The concentration of a large number of data points in low absolute relative error regions for training as well as test data sets is indicative of the superior prediction performance of the XGBoost algorithm. A majority of the data points exhibited relative errors less than 0.5% for all the three methods. XGBoost has the highest overall number of data points (1802) with absolute relative errors lower than 0.5%, which implies that it can predict about 98.3% of the entire (training + test) data set with greater than 99.5% relative accuracy. This is followed by Random Forest (1608 data points) and Extra Trees methods (1656 data points) corresponding to 87.67% and 90.29% of the data points with absolute relative errors less than 0.5%. As can be inferred from the R^2 plot (Fig. 2), the highest relative errors for the two temperatures (247.52 and 296.07 K belonging to the training and test data set respectively) correspond to the two points towards the right end of the histogram.

Next, we specifically illustrate the utility of the XGBoost model in predicting the inhibition influence of salts on hydrate formation. Figs. 4a-b show the experimentally obtained [64,68] pressure-temperature curves (non-solid symbols) associated with the formation of methane hydrates in the presence of NaCl and $CaCl_2$ respectively. It is seen that the addition of salts pushes the hydrate formation region to the left (higher pressures and/or lower temperatures needed for hydrate formation). Experimentally provided weight concentrations were used to estimate the activity

Table 3
Comparison between predictions (3 ML models) and experimental data [47] for hydrate formation under varying salt concentrations.

Gas composition (mol%)		Salt composition (wt%)			Experiments		Prediction			Relative error (%)		
CO_2	CH_4	NaCl	KCl	$CaCl_2$	P (MPa)	T (K)	RF	ET	XGBoost	RF	ET	XGBoost
15.3	84.7	0	0	0	3.41	277.56	277.77	277.51	277.54	0.076	−0.017	−0.007
16.7	83.3	0	0	0	5.14	281.50	280.83	280.72	281.49	−0.239	−0.279	−0.004
16.4	83.6	0	0	0	2.36	274.10	276.02	274.52	274.03	0.701	0.152	−0.025
17.9	82.1	0	0	0	7.53	284.84	284.63	283.93	284.72	−0.072	−0.321	−0.042
17.7	82.3	5.02	0	0	6.98	281.99	281.35	281.31	282.06	−0.227	−0.240	0.025
17.2	82.8	5.02	0	0	5.08	279.23	278.60	278.63	279.08	−0.226	−0.215	−0.055
16.1	83.9	5.02	0	0	3.26	274.98	273.01	274.15	273.50	−0.715	−0.302	−0.540
15.2	84.8	5.02	0	0	2.3	271.59	271.94	272.41	271.13	0.127	0.303	−0.171
17.4	82.6	9.99	0	0	6.56	278.98	277.91	278.55	279.05	−0.384	−0.153	0.023
16.1	83.9	9.99	0	0	3.1	272.07	271.19	272.07	271.95	−0.323	−0.001	−0.044
19.4	80.6	9.99	0	0	4.66	276.14	275.13	275.35	276.17	−0.366	−0.285	0.009
19.3	80.7	9.99	0	0	2.03	268.46	268.83	269.45	268.44	0.139	0.370	−0.007
18.9	81.1	15	0	0	7.31	277.17	275.54	275.71	277.24	−0.587	−0.527	0.025
19.4	80.6	15	0	0	2.88	269.12	268.84	269.34	268.90	−0.103	0.082	−0.082
18.2	81.8	15	0	0	7.37	276.79	276.06	276.55	276.92	−0.266	−0.086	0.046
19.8	80.2	15	0	0	4.4	273.14	272.18	272.76	273.23	−0.353	−0.138	0.034
17.3	82.7	15	0	0	1.86	264.78	266.22	266.27	264.71	0.544	0.564	−0.027
19.3	80.7	20	0	0	5.42	270.37	271.99	272.55	269.64	0.598	0.804	−0.271
19.1	80.9	20	0	0	2.12	261.95	264.03	263.21	262.16	0.796	0.481	0.079
19.8	80.2	20	0	0	3.86	267.48	267.72	267.51	267.32	0.089	0.010	−0.062
19.9	80.1	20.01	0	0	9.15	274.31	274.56	274.42	274.29	0.090	0.038	−0.007
19.7	80.3	0	5	0	2.04	271.38	272.14	271.68	271.49	0.278	0.112	0.042
19.8	80.2	0	5	0	4.46	278.83	278.04	278.32	278.60	−0.282	−0.185	−0.082
19.8	80.2	0	5	0	2.96	274.97	274.13	274.51	274.64	−0.306	−0.166	−0.122
19.8	80.2	0	5	0	6.43	281.98	281.39	281.46	281.97	−0.211	−0.183	−0.004
19.6	80.4	0	10	0	3.94	275.90	274.33	274.93	275.65	−0.569	−0.352	−0.091
19.7	80.3	0	10	0	1.83	269.16	269.48	269.86	269.25	0.118	0.259	0.034
19.8	80.2	0	10	0	2.59	272.16	271.29	271.98	272.17	−0.319	−0.065	0.005
19.8	80.2	0	10	0	5.66	278.96	277.52	278.56	278.92	−0.518	−0.142	−0.016
18.4	81.6	0	15	0	1.89	266.96	267.27	268.28	266.81	0.117	0.496	−0.055
19	81	0	15	0	2.62	270.07	269.34	270.23	270.17	−0.269	0.060	0.038
19.5	80.5	0	15	0	3.45	272.69	271.22	271.99	272.89	−0.539	−0.258	0.073
19.7	80.3	0	15	0	5.63	277.07	274.82	276.16	277.04	−0.812	−0.329	−0.012
18.7	81.3	0	0	9.91	1.96	268.59	269.95	270.02	268.77	0.506	0.532	0.068
19.7	80.3	0	0	9.91	4.32	276.04	275.95	275.89	275.84	−0.031	−0.054	−0.074
19.4	80.6	0	0	9.91	2.8	272.00	271.98	271.99	271.93	−0.008	−0.004	−0.027
19.7	80.3	0	0	9.91	6.08	279.13	279.36	278.91	279.67	0.082	−0.081	0.195
19	81	0	0	15	4.61	273.11	273.50	273.38	273.25	0.142	0.100	0.050
19.7	80.3	0	0	15	3	269.13	269.40	270.39	269.18	0.101	0.467	0.019
19.6	80.4	0	0	15	1.88	266.65	267.79	268.45	266.44	0.428	0.674	−0.080
19.7	80.3	0	0	15	7.09	276.88	276.77	278.17	277.88	−0.039	0.466	0.362
19.8	80.2	0	0	20	2.89	263.78	266.05	267.26	264.06	0.859	1.317	0.106
19.9	80.1	0	0	20	7.24	271.34	272.92	272.27	271.57	0.583	0.343	0.084
19.8	80.2	0	0	20	4.3	267.60	269.38	269.69	267.55	0.665	0.783	−0.019
19.9	80.1	0	0	20	9.46	273.69	274.96	274.12	273.80	0.464	0.157	0.041
19.4	80.6	5.01	0	10	5.04	273.67	273.88	274.08	273.66	0.075	0.151	−0.004
19.8	80.2	5.01	0	10	3.12	269.35	268.81	269.82	269.41	−0.199	0.175	0.023
19.7	80.3	5.01	0	10	2.06	265.54	267.34	269.10	266.40	0.677	1.341	0.324
19.9	80.1	5.01	0	10	7.14	276.71	275.35	276.75	276.30	−0.491	0.013	−0.147
19.8	80.2	0	10	5	1.65	265.07	266.17	266.91	265.14	0.413	0.693	0.028
19.9	80.1	0	10	5	4.6	274.76	273.73	274.22	274.66	−0.375	−0.197	−0.037
19.8	80.2	0	10	5	2.46	269.35	268.51	269.68	269.01	−0.314	0.123	−0.125
19.9	80.1	0	10	5	10.61	281.57	279.49	281.12	281.64	−0.738	−0.159	0.025
19.2	80.8	5	10	0	2.59	269.28	269.04	269.62	269.35	−0.088	0.128	0.025
19.8	80.2	5	10	0	10.41	281.27	278.34	279.26	281.21	−1.043	−0.716	−0.022
19.7	80.3	5	10	0	4.99	275.21	274.46	274.93	275.40	−0.273	−0.103	0.068
20.3	79.7	5	10	0	1.66	265.09	266.69	267.05	265.85	0.602	0.740	0.288
19.5	80.5	10.17	0	5.08	6.07	275.13	274.96	275.30	275.01	−0.061	0.062	−0.045
20.1	79.9	10.17	0	5.08	2.11	266.52	267.02	267.62	266.09	0.187	0.413	−0.160
20	80	10.17	0	5.08	3.1	269.18	269.06	269.52	269.07	−0.043	0.124	−0.042
19.4	80.6	10.17	0	5.08	9.71	278.74	277.23	278.84	278.34	−0.542	0.037	−0.142
48.6	51.4	10	0	10	3.53	268.08	269.15	268.52	268.36	0.398	0.165	0.105
49.7	50.3	10	0	10	2.15	264.14	268.70	269.48	266.13	1.725	2.020	0.753
49.4	50.6	10	0	10	5.16	270.80	271.38	271.15	271.11	0.214	0.128	0.114
49.7	50.3	10	0	10	3.52	268.21	269.39	269.58	268.97	0.439	0.511	0.283
46.5	53.5	10	5	0	4.59	275.48	273.58	275.22	275.41	−0.688	−0.093	−0.026
49.5	50.5	10	5	0	2.78	271.57	270.37	271.55	271.44	−0.440	−0.008	−0.047
49.8	50.2	10	5	0	1.82	268.04	268.07	268.55	267.69	0.012	0.190	−0.131
50	50	10	5	0	1.38	265.76	266.84	266.51	265.97	0.406	0.283	0.079

Table 4
Comparison between predictions (3 ML models) and experimental data [65] for hydrate formation from binary and ternary gas mixtures.

Gas composition (mol%)	Experiments		Prediction			Relative error (%)		
	P (MPa)	T (K)	RF	ET	XGBoost	RF	ET	XGBoost
CH_4: 89.26 N_2: 10.74	4.94	278.7	279.20	278.60	278.94	0.180	−0.035	0.086
	6.94	282.03	282.10	281.23	281.97	0.024	−0.284	−0.023
	10.40	285.64	285.85	285.03	285.78	0.072	−0.215	0.047
	14.98	288.68	288.53	288.38	288.49	−0.053	−0.103	−0.068
	20.02	290.97	291.01	290.61	290.87	0.015	−0.124	−0.035
	24.43	292.44	292.08	292.13	292.46	−0.123	−0.105	0.005
CH_4: 90.47 C_2H_6: 9.53	2.25	278.21	277.83	277.75	278.05	−0.138	−0.166	−0.059
	2.63	279.6	279.00	278.94	279.82	−0.213	−0.235	0.079
	4.19	283.69	283.07	282.72	283.44	−0.219	−0.342	−0.087
	7.21	288.12	287.39	287.28	288.05	−0.255	−0.290	−0.024
	9.99	290.44	290.11	290.31	290.41	−0.113	−0.046	−0.010
	14.85	292.97	292.41	292.34	292.81	−0.191	−0.214	−0.054
	19.89	294.63	294.26	294.56	294.55	−0.127	−0.024	−0.027
	23.20	295.52	294.65	294.51	295.54	−0.294	−0.343	0.008
CH_4: 97.07 C_3H_8: 2.93	2.64	247.52	262.33	260.32	250.28	5.985	5.171	1.115
	9.20	256.87	265.76	258.58	257.33	3.461	0.664	0.180
	14.97	259.33	265.20	260.14	258.59	2.264	0.314	−0.284
	19.81	260.59	264.96	260.59	260.07	1.675	0.000	−0.201
	24.36	261.53	268.53	261.53	261.96	2.675	0.000	0.163
CH_4: 94.97 N_2: 0.03	3.45	276.85	278.36	277.06	277.36	0.546	0.077	0.183
	4.87	279.95	280.24	279.85	279.87	0.103	−0.035	−0.028
	7.04	283.49	283.81	283.29	283.53	0.112	−0.072	0.012
	10.94	287.41	288.25	287.18	287.49	0.293	−0.081	0.027
	16.83	290.76	290.94	290.48	290.71	0.061	−0.098	−0.018
	23.98	293.41	292.85	293.41	293.24	−0.191	0.000	−0.057
C_2H_6: 85.15 C_3H_8: 14.85	0.83	276.66	276.99	276.32	276.88	0.117	−0.124	0.078
	1.09	278.92	278.70	278.90	278.91	−0.078	−0.008	−0.004
	1.51	281.49	281.36	281.43	281.51	−0.045	−0.020	0.006
	1.94	283.32	282.75	282.87	283.25	−0.203	−0.158	−0.026
	1.12	279.11	278.87	279.03	279.01	−0.086	−0.029	−0.037
	1.50	281.38	281.32	281.28	281.50	−0.020	−0.037	0.042
CH_4: 90.93 C_2H_6: 4.89	2.58	277.36	277.28	277.15	277.22	−0.028	−0.075	−0.050
	3.80	280.91	280.36	279.72	280.75	−0.195	−0.425	−0.057
	6.10	284.9	284.57	284.63	284.92	−0.118	−0.097	0.006
	8.36	287.45	287.14	287.22	287.34	−0.109	−0.082	−0.040
	13.81	290.93	290.59	290.62	290.90	−0.118	−0.108	−0.009
	18.82	292.82	292.77	292.56	292.78	−0.018	−0.089	−0.014
	23.83	294.23	293.72	294.23	294.19	−0.175	0.000	−0.014
CH_4: 84.52 C_2H_6: 12.55 C_3H_8: 2.93	1.20	277.1	277.21	277.09	276.97	0.041	−0.005	−0.046
	1.98	281.56	279.89	280.75	281.67	−0.593	−0.289	0.039
	4.00	287.42	285.91	287.02	287.58	−0.525	−0.138	0.056
	6.92	291.52	290.03	291.52	291.57	−0.511	0.000	0.017
	11.21	294.32	292.74	294.32	294.09	−0.538	0.000	−0.077
	17.03	296.14	294.68	296.14	296.51	−0.492	0.000	0.125
	24.47	298.14	295.99	298.14	298.00	−0.720	0.000	−0.047
CH_4: 95.02 C_2H_6: 3.98 C_3H_8: 1	2.16	279.1	275.34	276.36	275.67	−1.348	−0.981	−1.228
	3.99	284.3	282.22	282.91	284.01	−0.732	−0.489	−0.103
	6.96	288.56	286.96	288.23	288.50	−0.553	−0.115	−0.021
	10.94	291.6	290.08	291.54	291.54	−0.521	−0.021	−0.021
	17.36	294.04	292.34	294.04	294.15	−0.580	0.000	0.036
	24.15	295.76	293.86	295.76	295.80	−0.641	0.000	0.013
CO_2: 8.09 CH_4: 89.4 C_3H_8: 2.49 N_2: 0.02	1.69	279.19	277.49	278.44	278.97	−0.608	−0.269	−0.078
	3.47	285.09	282.34	283.85	285.02	−0.966	−0.435	−0.023
	6.91	290.16	287.30	289.48	290.21	−0.985	−0.234	0.018
	10.47	292.56	289.26	292.23	292.49	−1.129	−0.113	−0.025
	16.95	294.62	289.66	294.56	294.69	−1.684	−0.022	0.024
	23.57	296.07	283.04	281.87	289.47	−4.400	−4.796	−2.230
CO_2: 5.25 CH_4: 89.6 C_3H_8: 5.13 N_2: 0.02	2.96	279.01	278.55	278.26	278.95	−0.164	−0.268	−0.021
	5.00	283.54	282.80	282.83	283.35	−0.261	−0.252	−0.068
	7.74	287.08	286.91	286.98	286.99	−0.060	−0.036	−0.033
	11.88	290.04	289.83	289.86	290.04	−0.073	−0.061	−0.001
	17.52	292.28	292.37	290.82	293.02	0.030	−0.501	0.253
	24.33	294.21	293.86	294.21	294.37	−0.120	0.000	0.055

feature used in the ML models. From a chemical thermodynamics perspective, increased salt concentrations result in a reduction in activity, which will push the P-T curve to the left. Figs. 4a-b show the predictions of the XGBoost model; an excellent match is seen in the majority of the range of temperatures and pressures studied. There is a notable deviation for data points in lower ranges of temperatures (259.9–264 K) at 25.75 wt% $CaCl_2$ which can be attributed to a lack of experimental data points used for training the ML models at lower temperature ranges. Overall, the model can effectively capture the inhibition influence of salts containing monovalent as well as divalent cations in the solution.

Finally, we lay out the predicted HDTs (from Random Forest, Extra Trees and XGBoost algorithms) and the relative errors against experimental data reported in two detailed studies. The first study was by Dholabhai and Bishnoi [59], who studied hydrate formation in a mixture of CO_2 and CH_4 of varying compositions in the presence of various salts in water at varying weight percentages as outlined in Table 3. It is evident that while all the three ML models capture the hydrate forming conditions reasonably well (low error), the relative error exhibited by XGBoost is less than that of Random Forest (RF) and Extra Trees (ET) in most cases. The overall relative error predicted using XGBoost is predominantly less than 0.1% for most of the data points in Table 3. While the relative error of XGBoost is lesser than Random Forest and Extra Trees in a vast majority of the dataset, there are very few exceptions (CO_2/CH_4 at 16.1/83.9 mol% with 5.02 wt% NaCl; CO_2/CH_4 at 19.7/80.3 mol% with 9.91 wt% $CaCl_2$) where ET or RF exhibit slightly better performance. The highest relative errors occur for a gas composition consisting of CO_2/CH_4 at 49.7/50.3 mol% with NaCl and $CaCl_2$ at 10 wt% each.

The second study against which the ML based predictions are benchmarked is by Nixdorf and Oellrich [77] who studied hydrate dissociation conditions for various binary and ternary gas mixtures as outlined in Table 4. The highest relative errors are observed for data points containing CH_4/C_3H_8 at 97.07/2.93 mol% respectively. One of the primary reasons underlying the poor prediction performance is the lack of data points at lower temperatures (247–261 K) over which these data points have been evaluated. Furthermore, the relative error increases slightly for predictions by Random Forest model when a third gas component is introduced.

5. Conclusions

In summary, a set of thermodynamic features such as partial pressures, specific gravity of hydrate precursor mixtures and the activity contribution due to salts were used in three ML-based models to predict HDTs. Using such physics-based features in ML frameworks enables the models to track individual contributions of gases towards hydrate formation. Importantly, it extends the utility of the model to predict the inhibition effect for any other salt via calculation of an activity term. An extensive databank comprising of more than 1800 experimental data points was employed to train and evaluate the prediction accuracies of the ML models. While the use of Random Forest (RF) and Extra Trees (ET) has been reported previously, we also use extreme gradient boosting (XGBoost) to predict hydrate equilibria and report a considerable improvement in prediction accuracies compared to RF and ET. The overall coefficient of determination (R^2) percentage is greater than 97% and the overall average absolute relative deviation (AARD) is lower than 0.25% respectively for all three models. XGBoost exhibits the highest R^2 accuracy and the lowest AARD for training and test data, which highlights its superior capabilities to predict hydrate equilibrium conditions.

Finally, it is important to note that the performance of ML models used in the present study is determined by the experimental dataset employed for training the algorithms. Non-uniformities or gaps in the distribution of experimental data over a certain range of values, or thermodynamically inconsistent data affect the quality of predictions as is the case for extreme temperature and pressure in the case of hydrates. In the present work we have used an exhaustive data bank (>1800 data points) so that an error or discrepancy in reported experimental data would not significantly bias the model. Future efforts should involve application of a consistent screening criteria to assess the thermodynamic consistency of the experimental data employed [89] to train such algorithms to further improve the performance and prediction accuracies of the developed ML models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Palash V. Acharya: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft. **Vaibhav Bahadur:** Conceptualization, Resources, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgements

Authors acknowledge NSF-CBET grant 1653412 and a grant from the office of Vice President of Research at UT Austin for partial support of this work. Palash V. Acharya acknowledges the support of the University Graduate Continuing Fellowship from UT Austin.

References

[1] EIAInternational Energy Outlook, Washington, D.C., 2019.
[2] V. Balzani, N. Armaroli, Energy for a Sustainable World: from the Oil Age to a Sun-Powered Future, John Wiley & Sons, 2010.
[3] E. Dendy Sloan, C. Koh, Clathrate Hydrates of Natural Gases, third ed., 2007, doi:10.1201/9781420008494.
[4] W.-.H. Ip, Advances in Geosciences, World Scientific Publishing Company, 2010, doi:10.1142/7158-vol16.
[5] Z.W. Ma, P. Zhang, H.S. Bao, S. Deng, Review of fundamental properties of CO2hydrates and CO2capture and separation using hydration method, Renew. Sustain. Energy Rev. 53 (2016) 1273–1302, doi:10.1016/j.rser.2015.09.076.
[6] P. Babu, P. Linga, R. Kumar, P. Englezos, A review of the hydrate based gas separation (HBGS) process forcarbon dioxide pre-combustion capture, Energy 85 (2015) 261–279, doi:10.1016/j.energy.2015.03.103.
[7] P.V. Acharya, A. Kar, A. Shahriari, A. Bhati, A. Mhadeshwar, V. Bahadur, Aluminum-based promotion of nucleation of carbon dioxide hydrates, J. Phys. Chem. Lett. 11 (2020) 1477–1482, doi:10.1021/acs.jpclett.9b03485.
[8] I. Chatti, A. Delahaye, L. Fournaison, J.P. Petitet, Benefits and drawbacks of clathrate hydrates: a review of their areas of interest, Energy Convers. Manag. 46 (2005) 1333–1343, doi:10.1016/j.enconman.2004.06.032.
[9] J. Javanmardi, M. Moshfeghian, Energy consumption and economic evaluation of water desalination by hydrate phenomenon, Appl. Therm. Eng. 23 (2003) 845–857, doi:10.1016/S1359-4311(03)00023-1.
[10] D.B. Carson, D.L. Katz, Natural gas hydrates, Trans. AIME 146 (01) (1942) 150–158.
[11] D.L. Katz, Prediction of conditions for hydrate formation in natural gases, Trans. AIME 160 (1945) 140–149, doi:10.2118/945140-g.
[12] J. Loh, R.N. Maddox, J.H. Erbar, New hydrate-formation data reveal differences, OIL GAS J. 81 (1983) 96–98.
[13] E.D. SLOAN, Phase equilibria of natural gas hydrates, (1984) 163–169.
[14] B.K. Berge, Hydrate predictions on a microcomputer, Pet. Ind. Appl. Microcomput., Society of Petroleum Engineers, 1986, doi:10.2118/15306-MS.
[15] R. Kobayashi, K.Y. Song, E.D. Sloan, Phase behavior of water/hydrocarbon systems, Pet. Eng. Handb., 1987.
[16] A. Bahadori, H.B. Vuthaluru, A novel correlation for estimation of hydrate forming condition of natural gases, J. Nat. Gas Chem. 18 (2009) 453–457, doi:10.1016/S1003-9953(08)60143-7.
[17] M.M. Ghiasi, Initial estimation of hydrate formation temperature of sweet natural gases based on new empirical correlation, J. Nat. Gas Chem. 21 (2012) 508–512, doi:10.1016/S1003-9953(11)60398-8.
[18] G.D. Holder, S.P. Zetts, N. Pradhan, Phase behavior in systems containing clathrate hydrates: a review, Rev. Chem. Eng. 5 (1988).
[19] J.H. van der Waals, J.C. Platteeuw, in: Clathrate Solutions, John Wiley & Sons, Ltd, 2007, pp. 1–57, doi:10.1002/9780470143483.ch1.
[20] W.R. Parrish, J.M. Prausnitz, Dissociation pressures of gas hydrates formed by gas mixtures, Ind. Eng. Chem. Process Des. Dev. 11 (1972) 26–35, doi:10.1021/i260041a006.
[21] J. Javanmardi, M. Moshfeghian, R.N. Maddox, Simple method for predicting gas-hydrate-forming conditions in aqueous mixed-electrolyte solutions, Energy Fuels 12 (1998) 219–222, doi:10.1021/ef9701652.
[22] J. Javanmardi, M. Moshfeghian, R.N. Maddox, An accurate model for prediction of gas hydrate formation conditions in mixtures of aqueous electrolyte solutions and alcohol, Can. J. Chem. Eng. 79 (2001) 367–373, doi:10.1002/cjce.5450790309.
[23] D. Shabani, M.M. Rashtchian, C. Ghotbi, V. Taghikhani, G. KHAYAT, Prediction of hydrate formation for the systems containing single and mixed electrolyte solutions, Iran, J. Chem. Chem. Eng. 26 (2007) 35–45.

[24] H. Tavasoli, F. Feyzi, M.R. Dehghani, F. Alavi, Prediction of gas hydrate formation condition in the presence of thermodynamic inhibitors with the Elliott–Suresh–Donohue Equation of State, J. Pet. Sci. Eng. 77 (2011) 93–103, doi:10.1016/j.petrol.2011.02.002.

[25] P.F. Ferrari, A.Z. Guembaroski, M.A. Marcelino Neto, R.E.M. Morales, A.K. Sum, Experimental measurements and modelling of carbon dioxide hydrate phase equilibrium with and without ethanol, Fluid Phase Equilib. 413 (2016) 176–183, doi:10.1016/j.fluid.2015.10.008.

[26] H. Delavar, A. Haghtalab, Thermodynamic modeling of gas hydrate formation conditions in the presence of organic inhibitors, salts and their mixtures using UNIQUAC model, Fluid Phase Equilib. 394 (2015) 101–117, doi:10.1016/j.fluid.2015.03.008.

[27] T.H. Sirino, M.A. Marcelino Neto, D. Bertoldi, R.E.M. Morales, A.K. Sum, Multiphase flash calculations for gas hydrates systems, Fluid Phase Equilib. 475 (2018) 45–63, doi:10.1016/j.fluid.2018.07.029.

[28] H. Najibi, A. Chapoy, H. Haghighi, B. Tohidi, Experimental determination and prediction of methane hydrate stability in alcohols and electrolyte solutions, Fluid Phase Equilib. 275 (2009) 127–131, doi:10.1016/j.fluid.2008.09.020.

[29] M.A. Mahabadian, A. Chapoy, R. Burgass, B. Tohidi, Development of a multiphase flash in presence of hydrates: experimental measurements and validation with the CPA equation of state, Fluid Phase Equilib. 414 (2016) 117–132, doi:10.1016/j.fluid.2016.01.009.

[30] S. Shahnazar, N. Hasan, Gas hydrate formation condition: review on experimental and modeling approaches, Fluid Phase Equilib. 379 (2014) 72–85, doi:10.1016/j.fluid.2014.07.012.

[31] I.A. Basheer, M. Hajmeer, Artificial neural networks: fundamentals, computing, design, and application, J. Microbiol. Methods 43 (2000) 3–31, doi:10.1016/S0167-7012(00)00201-3.

[32] M. Paliwal, U.A. Kumar, Neural networks and statistical techniques: a review of applications, Expert Syst. Appl. 36 (2009) 2–17, doi:10.1016/j.eswa.2007.10.005.

[33] K.P. Ferentinos, Biological engineering applications of feedforward neural networks designed and parameterized by genetic algorithms, Neural Networks 18 (2005) 934–950, doi:10.1016/j.neunet.2005.03.010.

[34] A.A. Elgibaly, A.M. Elkamel, A new correlation for predicting hydrate formation conditions for various gas mixtures and inhibitors, Fluid Phase Equilib. 152 (1998) 23–42, doi:10.1016/S0378-3812(98)00368-9.

[35] A. Elgibaly, A. Elkamel, Optimal hydrate inhibition policies with the aid of neural networks, Energy Fuels 13 (1999) 105–113, doi:10.1021/ef980129i.

[36] G. Zahedi, Z. Karami, H. Yaghoobi, Prediction of hydrate formation temperature by both statistical models and artificial neural network approaches, Energy Convers. Manag. 50 (2009) 2052–2059, doi:10.1016/j.enconman.2009.04.005.

[37] A. Chapoy, A.H. Mohammadi, D. Richon, Predicting the hydrate stability zones of natural gases using artificial neural networks, Oil Gas Sci. Technol.–Rev. l'IFP. 62 (2007) 701–706, doi:10.2516/ogst:2007048.

[38] M. Ghavipour, M. Ghavipour, M. Chitsazan, S.H. Najibi, S.S. Ghidary, Experimental study of natural gas hydrates and a novel use of neural network to predict hydrate formation conditions, Chem. Eng. Res. Des. 91 (2013) 264–273, doi:10.1016/j.cherd.2012.08.010.

[39] Y. Song, H. Zhou, P. Wang, M. Yang, Prediction of clathrate hydrate phase equilibria using gradient boosted regression trees and deep neural networks, J. Chem. Thermodyn. 135 (2019) 86–96, doi:10.1016/j.jct.2019.03.030.

[40] H. Yarveicy, M.M. Ghiasi, Modeling of gas hydrate phase equilibria: extremely randomized trees and LSSVM approaches, J. Mol. Liq. 243 (2017) 533–541, doi:10.1016/j.molliq.2017.08.053.

[41] M. Mesbah, E. Soroush, M. Rezakazemi, Development of a least squares support vector machine model for prediction of natural gas hydrate formation temperature, Chinese J. Chem. Eng. 25 (2017) 1238–1248, doi:10.1016/j.cjche.2016.09.007.

[42] M.M. Ghiasi, H. Yarveicy, M. Arabloo, A.H. Mohammadi, R.M. Behbahani, Modeling of stability conditions of natural gas clathrate hydrates using least squares support vector machine approach, J. Mol. Liq. 223 (2016) 1081–1092, doi:10.1016/j.molliq.2016.09.009.

[43] M.K.B. Landgrebe, D. Nkazi, Toward a Robust, universal predictor of gas hydrate equilibria by means of a deep learning regression, ACS Omega 4 (2019) 22399–22417, doi:10.1021/acsomega.9b02961.

[44] N. Rebai, A. Hadjadj, A. Benmounah, A.S. Berrouk, S.M. Boualleg, Prediction of natural gas hydrates formation using a combination of thermodynamic and neural network modeling, J. Pet. Sci. Eng. 182 (2019) 106270, doi:10.1016/j.petrol.2019.106270.

[45] A. Hosseinzadeh, A. Mohammadi, E. Soroush, Predicting semiclathrate hydrates dissociation pressure using a rigorous machine learning approach, J. Dispers. Sci. Technol. 41 (2020) 863–872, doi:10.1080/01932691.2019.1614028.

[46] A.H. Mohammadi, D. Richon, Determination of gas hydrate safety margin using specific gravity data of salt or organic inhibitor aqueous solution, Ind. Eng. Chem. Res. 46 (2007) 3852–3857, doi:10.1021/ie060908j.

[47] A.H. Mohammadi, J.F. Martínez-López, D. Richon, Determination of hydrate stability zone using electrical conductivity data of salt aqueous solution, Fluid Phase Equilib. 253 (2007) 36–41, doi:10.1016/j.fluid.2007.01.006.

[48] S.M. Lundberg, G.G. Erion, S.-I. Lee, Consistent individualized feature attribution for tree ensembles, ArXiv Prepr. ArXiv1802.03888. (2018).

[49] W. Zeng, A. Davoodi, R.O. Topaloglu, Explainable DRC hotspot prediction with random forest and SHAP tree explainer, in: 2020 Des. Autom. Test Eur. Conf. Exhib., IEEE, 2020, pp. 1151–1156, doi:10.23919/DATE48585.2020.9116488.

[50] X. Zhao, Z. Qiu, Z. Zhang, Y. Zhang, Relationship between the gas hydrate suppression temperature and water activity in the presence of thermodynamic hydrate inhibitor, Fuel 264 (2019) 116776, doi:10.1016/j.fuel.2019.116776.

[51] J. Javanmardi, M. Moshfeghian, R.N. Maddox, An accurate model for prediction of gas hydrate formation conditions in mixtures of aqueous electrolyte solutions and alcohol, Can. J. Chem. Eng. 79 (2001) 367–373, doi:10.1002/cjce.5450790309.

[52] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, doi:10.1023/A:1010933404324.

[53] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (2006) 3–42, doi:10.1007/s10994-006-6226-1.

[54] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., ACM, New York, NY, USA, 2016, pp. 785–794, doi:10.1145/2939672.2939785.

[55] L. Breiman, Bagging predictors, Mach. Learn. 24 (1996) 123–140, doi:10.1007/bf00058655.

[56] T.K. Ho, The random subspace method for constructing decision forests, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 832–844, doi:10.1109/34.709601.

[57] N. Dholabhai, P.D. Kalogerakis, P.R. Bishnoi, Equilibrium conditions for carbon dioxide hydrate formation in aqueous electrolyte solutions, J. Chem. Eng. Data 38 (1993) 650–654.

[58] P. Englezos, Y.T. Ngan, Incipient equilibrium data for propane hydrate formation in aqueous solutions of sodium chloride, potassium chloride and calcium chloride., J. Chem. Eng. Data 38 (1993) 250–253.

[59] P.D. Dholabhai, P.R. Bishnoi, Hydrate equilibrium conditions in aqueous electrolyte solutions: mixtures of methane and carbon dioxide, J. Chem. Eng. Data 39 (1994) 191–194.

[60] D.H. Mei, J. Liao, J.T. Yang, T.M. Guo, Experimental and modeling studies on the hydrate formation of a methane + nitrogen gas mixture in the presence of aqueous electrolyte solutions, Ind. Eng. Chem. Res. 35 (1996) 4342–4347, doi:10.1021/ie9601662.

[61] S.P. Kang, M.K. Chun, H. Lee, Phase equilibria of methane and carbon dioxide hydrates in the aqueous $MgCl_2$ solutions, Fluid Phase Equilib. 147 (1998) 229–238, doi:10.1016/s0378-3812(98)00233-7.

[62] T. Maekawa, Equilibrium conditions for gas hydrates of methane and ethane mixtures in pure water and sodium chloride solution, Geochem. J. 35 (2001) 59–66, doi:10.2343/geochemj.35.59.

[63] M.D. Jager, E.D. Sloan, The effect of pressure on methane hydration in pure water and sodium chloride solutions, Fluid Phase Equilib. 185 (2001) 89–99, doi:10.1016/S0378-3812(01)00459-9.

[64] M. Kharrat, D. Dalmazzone, Experimental determination of stability conditions of methane hydrate in aqueous calcium chloride solutions using high pressure differential scanning calorimetry, J. Chem. Thermodyn. 35 (2003) 1489–1505, doi:10.1016/S0021-9614(03)00121-6.

[65] Z. Atik, C. Windmeier, L.R. Oellrich, Experimental gas hydrate dissociation pressures for pure methane in aqueous solutions of $MgCl_2$ and $CaCl_2$ and for a (methane + ethane) gas mixture in an aqueous solution of (NaCl + $MgCl_2$), J. Chem. Eng. Data 51 (2006) 1862–1867, doi:10.1021/je060225a.

[66] A.H. Mohammadi, W. Afzal, D. Richon, Gas hydrates of methane, ethane, propane, and carbon dioxide in the presence of single NaCl, KCl, and $CaCl_2$ aqueous solutions: experimental measurements and predictions of dissociation conditions, J. Chem. Thermodyn. 40 (2008) 1693–1697, doi:10.1016/j.jct.2008.06.015.

[67] H. Haghighi, A. Chapoy, B. Tohidi, 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.–Rev. l'IFP. 64 (2009) 141–154, doi:10.2516/ogst:2008043.

[68] J.L. De Roo, C.J. Peters, R.N. Lichtenthaler, G.A.M. Diepen, Occurrence of methane hydrate in saturated and unsaturated solutions of sodium chloride and water in dependence of temperature and pressure, AIChE J. 29 (1983) 651–657, doi:10.1002/aic.690290420.

[69] T. Maekawa, S. Itoh, S. Sakata, S. Igari, N. Imai, Pressure and temperature conditions for methane hydrate dissociation in sodium chloride solutions, Geochem. J. 29 (1995) 325–329, doi:10.2343/geochemj.29.325.

[70] P.D. Dholabhai, P. Englezos, N. Kalogerakis, P.R. Bishnoi, Equilibrium conditions for methane hydrate formation in aqueous mixed electrolyte solutions, Can. J. Chem. Eng. 69 (1991) 800–805, doi:10.1002/cjce.5450690324.

[71] R. Nakane, E. Gima, R. Ohmura, I. Senaha, K. Yasuda, Phase equilibrium condition measurements in carbon dioxide hydrate forming system coexisting with sodium chloride aqueous solutions, J. Chem. Thermodyn. 130 (2019) 192–197, doi:10.1016/j.jct.2018.10.008.

[72] B. Tohidi, R.W. Burgass, A. Danesh, A.C. Todd, Hydrate inhibition effect of produced water: part 1—ethane and propane simple gas hydrates, Offshore Eur., Society of Petroleum Engineers, 1993, doi:10.2118/26701-MS.

[73] G.D. Holder, G.C. Grigoriou, Hydrate dissociation pressures of (methane + ethane + water) existence of a locus of minimum pressures, J. Chem. Thermodyn. 12 (1980) 1093–1104, doi:10.1016/0021-9614(80)90166-4.

[74] S.G. Paranjpe, S.L. Patil, V.A. Kamath, S.P. Godbole, Hydrate equilibria for binary and ternary mixtures of methane, propane, isobutane, and n-butane. Effect of salinity, SPE Reserv. Eng. (Society Pet. Eng. 4 (1989) 446–454 16871, doi:10.2118/16871-pa.

[75] S. Adisasmito, R.J. Frank, E.D. Sloan, Hydrates of carbon dioxide and methane mixtures, J. Chem. Eng. Data 36 (1991) 68–71, doi:10.1021/je00001a020.

[76] S. Adisasmito, E.D. Sloan Jr, Hydrates of hydrocarbon gases containing carbon dioxide, J. Chem. Eng. Data 37 (1992) 343–349.

[77] J. Nixdorf, L.R. Oellrich, Experimental determination of hydrate equilibrium conditions for pure gases, binary and ternary mixtures and natural gases, Fluid Phase Equilib. 139 (1997) 325–333, doi:10.1016/s0378-3812(97)00141-6.

[78] H.-J. Ng, D.B. Robinson, The role ofn-butane in hydrate formation, AIChE J. 22 (1976) 656–661, doi:10.1002/aic.690220404.

[79] G.D. Holder, J.H. Hand, Multiple-phase equilibria in hydrates from methane, ethane, propane and water mixtures, AIChE J. 28 (1982) 440–447, doi:10.1002/aic.690280312.

[80] J. Jhaveri, D.B. Robinson, Hydrates in the methane-nitrogen system, Can. J. Chem. Eng. 43 (1965) 75–78, doi:10.1002/cjce.5450430207.

[81] B.J. Wu, D.B. Robinson, H.J. Ng, Three- and four-phase hydrate forming conditions in methane + isobutane + water, J. Chem. Thermodyn. 8 (1976) 461–469, doi:10.1016/0021-9614(76)90067-7.

[82] W. Deaton, E. Frost, Gas Hydrates and their Relation to the Operation of Natural-Gas Pipe Lines, American Gas Association, 1949.

[83] V.K. Verma, Gas Hydrates from Liquid Hydrocarbon-Water Systems, University of Michigan, 1974.

[84] C.H. Unruh, D.L. Katz, Gas hydrates of carbon dioxide-methane mixtures, J. Pet. Technol. 1 (1949) 83–86, doi:10.2118/949983-g.

[85] K. Ohgaki, K. Takano, H. Sangawa, T. Matsubara, S. Nakano, Methane exploitation by carbon dioxide from gas hydrates. Phase equilibria for CO_2-CH_4 mixed hydrate system., J. Chem. Eng. Japan 29 (1996) 478–483, doi:10.1252/jcej.29.478.

[86] S.S. Fan, T.M. Guo, Hydrate formation of CO_2-rich binary and quaternary gas mixtures in aqueous sodium chloride solutions, J. Chem. Eng. Data 44 (1999) 829–832, doi:10.1021/je990011b.

[87] Y.T. Seo, H. Lee, J.H. Yoon, Hydrate phase equilibria of the carbon dioxide, methane, and water system, J. Chem. Eng. Data 46 (2001) 381–384, doi:10.1021/je000237a.

[88] H.J. Ng, J.P. Petrunia, D.B. Robinson, Experimental measurement and prediction of hydrate forming conditions in the nitrogen-propane-water system, Fluid Phase Equilib. 1 (1977) 283–291, doi:10.1016/0378-3812(77)80011-3.

[89] K.S. Pitzer, Electrolytes. From dilute solutions to fused salts, J. Am. Chem. Soc. 102 (1980) 2902–2906, doi:10.1021/ja00529a006.

[90] A. Haghtalab, A. Shojaeian, S.H. Mazloumi, Nonelectrolyte NRTL-NRF model to study thermodynamics of strong and weak electrolyte solutions, J. Chem. Thermodyn. 43 (2011) 354–363, doi:10.1016/j.jct.2010.10.004.

[91] S. Li, Y. Li, J. Wang, K. Ge, L. Yang, Prediction of gas hydrate formation conditions in the presence of electrolytes using an N-NRTL-NRF activity coefficient model, Ind. Eng. Chem. Res. 59 (2020) 6269–6278, doi:10.1021/acs.iecr.9b06411.

[92] K.S. Pitzer, J.M. Simonson, Thermodynamics of multicomponent, miscible, ionic systems: theory and equations, J. Phys. Chem. 90 (1986) 3005–3009, doi:10.1021/j100404a042.

[93] M. Margules, Uber die Zusammensetzung der gesattigten Dampfe von Mischungen, 104, Sitzungsber Akad Wiss, Wien, 1895, pp. 1243–1278.

[94] A. Pedregosa, F. Varoquaux, G. Gramfort, V. Michel, B. Thirion, O. Grisel, P. Blondel, M. Prettenhofer, R. Weiss, J. Dubourg, V. Vanderplas, A. Passos, M. Cournapeau, D. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in {P}ython, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[95] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Springer, New York, 2013.