Semarak Engineering Journal

Journal homepage:
https://semarakilmu.my/index.php/sej/index
ISSN: 3036-0145

Novel Hyperparameter Optimization for an SVM-based CO_2/CH_4 Hydrate Equilibrium Conditions for Ammonium-based Ionic Liquids

Dileep Kumar[1,2], Muhammad Saad Khan[3], Bhajan Lal[1,2,*]

[1] Chemical Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 32610, Perak, Malaysia
[2] Centre for Sustainable Energy, Environment and Social Innovation (SEESI), Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 32610, Perak, Malaysia
[3] Interdisciplinary Research Center of Membrane and Water Security, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

ARTICLE INFO

Article history:
Received 10 January 2026
Received in revised form 4 March 2026
Accepted 31 March 2026
Available online 11 May 2026

Keywords:
Ammonium-based ionic liquids; hydrate inhibition; hyperparameter optimization; support vector machine model

ABSTRACT

Gas hydrates are reservoirs of CO_2, CH_4, and other gases, found in seabed subsurface sediments, in permafrost regions and in deep freshwater lakes. These gases can be transported using hydrate technologies. However, when gas is transported through oil and gas pipelines, the formation of gas hydrates inside the pipelines causes blockage of the pipelines, damages the pipelines, and sometimes causes loss of life. Several hydrate mitigation methods exist, and chemical inhibition has gained significant attention. Ammonium-based ionic liquids have received considerable attention as chemical inhibitors because of their eco-friendliness, low volatility, and reusability. However, these experimental methods are time-consuming and expensive. Hence, machine learning model building is the best and most suitable complement to experimental work for gas hydrate inhibition. A support vector machine (SVM) learning model was built with and without hyperparameter optimization for CO_2/CH_4 gas hydrate inhibition using ammonium-based ionic liquids (AILs). The data were collected from the literature and thermodynamic models. Outliers were removed from the data. The models were trained and tested using a 70:30 ratio. The model performance was analyzed using experimental versus predicted temperature, residual analysis, cumulative probability plot, and error analysis metrics and unseen data. It was concluded that SVM model hyperparameter optimization is a necessary for CO_2/CH_4 hydrate inhibition using AILs.

1. Introduction

Gas hydrates are crystalline compounds resembling ice, which form when water and gas molecules come together under high-pressure and low-temperature conditions. In these structures, gas molecules, referred to as guests, are trapped within cavities created by hydrogen-bonded water molecules known as hosts [1]. The occurrence of gas hydrates in the oil and gas sector poses a challenge to flow assurance because it can lead to the obstruction of pipelines and disruption of

---

[*] *Corresponding author.*
E-mail address: bhajan.lal@utp.edu.my

https://doi.org/10.37934/sej.14.1.103112

operations [2]. Consequently, the industry is keen on developing cost-effective and eco-friendly strategies to prevent gas hydrate formation. Several approaches have been suggested to mitigate gas hydrate formation, including chemical inhibition, depressurization, thermal heating, and dehydration [3]. Among these, chemical inhibition is the most widely adopted method [4]. This method involves introducing chemicals to restrain gas flow outside the hydrate-formation zones. These chemicals, known as chemical inhibitors, either modify the phase behavior of gas hydrates or postpone their formation. There are two primary categories of chemical inhibitors, kinetic hydrate inhibitors (KHIs) and thermodynamic hydrate inhibitors (THIs). KHIs delay hydrate formation, whereas THIs adjust the hydrate equilibrium curve to lower temperatures and higher pressures [5].

Ionic liquids have attracted significant interest as innovative gas hydrate inhibitors because of their eco-friendly characteristics and state as molten salts at ambient temperature. Their distinctive features, including customizable cations and anions, very low vapor pressures, and straightforward synthesis from relatively low-cost materials, have contributed to their appeal. Additionally, they carry electrostatic charges and can form hydrogen bonds with water, making them promising candidates for effective and efficient THIs [6]. Research indicates that ILs can be reused after application, which supports their potential use in the oil and gas sectors [7]. Various experimental studies have been conducted on hydrate inhibition using ILs [8]. However, experimental studies are time-consuming, and there are chemical and gas purchasing issues; hence, machine learning modelling is a viable option.

Machine learning modelling has gained considerable attention in gas hydrates and various dynamic fields of research. Artificial Neural Networks and least squares support vector machine (LSSVM) methods have been utilized to model the hydrate behavior of natural gas mixtures, with the aim of understanding hydrate formation characteristics [9,10]. The LSSVM approach was used to forecast the hydrate phase boundary conditions for CO_2, CH_4, H_2, N_2, and water-soluble organic promoters [11]. Furthermore, Yarveicy and Ghiasi [12] built an LSSVM and extremely randomized tree (extra tree) models for predicting the hydrate formation/dissociation conditions of various mixtures of gases with pure water, salt, and alcohol solutions. The authors found that extra tree performs better than the LSSVM model. Xu et al. [13] implemented and compared five machine learning algorithms to predict the formation temperature of methane hydrate under saltwater conditions: Multiple Linear Regression, Gradient Boosting Regression, Random Forest, support vector machine (SVM), and k-nearest Neighbor. However, there is limited research available on SVM learning model hyperparameter optimization for CO_2/CH_4 hydrate inhibition using ammonium-based ionic liquids (AILs). Hence, the main objective of this study is to develop an SVM model with and without hyperparameter optimization, and to analyze its performance.

2. Methodology

2.1 Data Collection

Data on the composition of the CO_2/CH_4 mixture, phase equilibrium temperature, pressure, and ionic liquid concentration were obtained from literature [14–17] and calculated using the Dickens and Quinby-Hunt model [18]. In total, 981 data points were collected. The ranges of the data are listed in Table 1.

Table 1
The components of the dataset

Components	Range
CO_2 Composition (%)	(0-100)
CH_4 Composition (%)	(0-100)
CO_2 Molar Mass (g/mol)	(0-44.01)
CH_4 Molar Mass (g/mol)	(0-16.04)
CO_2/CH_4 Molar Mass (g/mol)	(16.04-44.01)
CO_2/CH_4 Gas gravity	(0.55-1.52)
Temperature (K)	(263-288.9)
Pressure (MPa)	(1.84-14.37)
Tetramethylammonium hydroxide (wt%)	(0-30)
Tetramethylammonium chloride (wt%)	(0-30)
Tetraethylammonium hydroxide (wt%)	(0-30)
Tetrapropylammonium hydroxide (wt%)	(0-30)
Tetrabutylammonium hydroxide (wt%)	(0-30)

2.2 Outlier Analysis

Outlier analysis is a critical process used to identify individual data points or groups of data that deviate significantly from the bulk of a population within a dataset [19]. In the context of sources, this analysis is necessary for developing robust predictive models because outliers, often caused by experimental errors, can introduce uncertainties and lower the overall accuracy of the results. The graphical tool, Williams plot, was used for outlier detection [20].

2.3 SVM Model

An SVM is a supervised machine learning methodology based on the Vapnik-Chervonenkis theory and Structural Risk Minimization (SRM) principle [21,22]. Although originally designed for classification, it is widely applied to regression problems, often called Support Vector Regression (SVR), to predict the thermodynamic stability and inhibition conditions of natural gas hydrates [23]. The SVM model functions by mapping nonlinear input data into a high-dimensional feature space. Within this space, the disorganized data points become linearly separable, allowing the model to calculate a linear regression function that minimizes the training error [24]. To handle nonlinear relationships, SVM uses a kernel function to reorganize the original input space. Common kernels include linear, polynomial, sigmoid, and radial basis functions (RBF). The Radial Basis Function (RBF), is defined as [20]:

$$G\left(x_i, x_j\right) = \exp\left(-\frac{\left\| x - x_i \right\|^2}{\sigma^2}\right)$$

In this function, s^2 represents kernel bandwidth. The terms x, x_i, and x_j are support vectors that represent the independent input variables. G is the RBF kernel function.

2.4 Hyperparameter Optimization

In this modeling work, the hyperparameter optimization was conducted using the regularization factor 'C', gamma 'g', and epsilon 'Œ'. The regularization factor controls the trade-off between the model complexity and the degree to which training errors are permitted. Gamma determines the reach or "influence" of individual training points on the resulting decision surface, and epsilon represents the acceptable tolerance or accuracy indicator within the system.

2.5 Statistical Parameters

The statistical parameters, Correlation Coefficient (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) were used to evaluate the accuracy, robustness, and generalization ability of the mathematical models. R^2 represents the degree of agreement between the values predicted by the model and the actual experimental data. An R^2 value close to 1.0 signifies strong prediction accuracy. The MAE and RMSE were used to measure the average magnitude of the error between the predicted and experimental values. MAE and RMSE values close to zero indicate a strong prediction accuracy [13,25].

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(X_{pre} - X_{\exp}\right)}{\sum_{i=1}^{N} \left(X_{pre} - \bar{X}_{\exp}\right)}, \quad i = 1,2,3,...N$$

$$MAE = \frac{\sum_{i=1}^{N} \left|X_{\exp} - X_{pre}\right|}{N}, \quad i = 1,2,3,...N$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(X_{\exp} - X_{pre}\right)^2}{N}}, \quad i = 1,2,3,...N$$

Where X_{pre} is Predicted Temperature, $X_{\exp}$ is the experimental temperature, $\bar{X}_{\exp}$ the average experimental temperature. N is the number of data points.

2.6 SVM Model Building

The SVM model for CO_2/CH_4 hydrate inhibition using AILs was built with and without hyperparameter optimization. The hyperparameters selected for the model are C =1000, g =0.01 and Œ=0.05. The 981 data points were split into a training dataset and a testing dataset at a 70:30 ratio. An outlier analysis of the dataset was conducted. The with and without optimization model was trained and tested using data. The performance of the model was evaluated using an experimental versus predicted plot, residual plot, cumulative probability plot, and error analysis metrics. In addition, the performance of the model was analyzed using unseen data. The unseen data was calculated using the Dickens and Quinby-Hunt model.

3. Results and Discussion

3.1 Outlier detection

An outlier analysis was conducted using William's plot of the dataset containing 981 data points, as shown in the Figure 1. Only one outlier was detected in the analysis. This clearly indicates that the dataset was highly reliable for model training and testing.

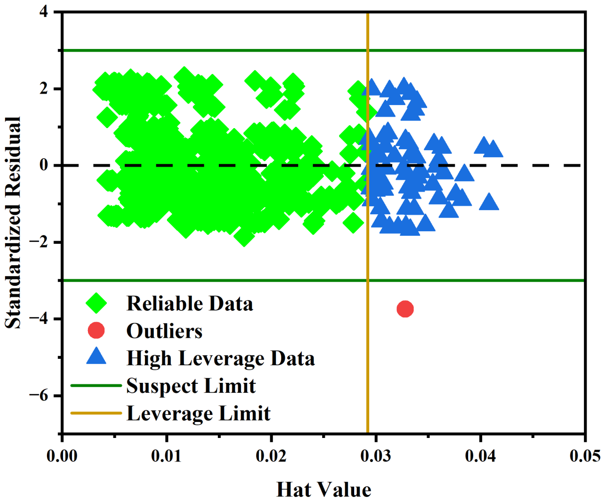

Fig. 1. Outlier detection of the dataset

3.2 Experimental versus Predicted Temperature

The experimental versus predicted temperatures for CO_2/CH_4 hydrate inhibition using ILs are shown in Figure 2(a) and Figure 2(b) with and without optimization. It was found that without optimization, the experimental versus predicted temperature followed the diagonal line; however, many data points of the training and testing datasets did not follow the diagonal line. A few of the data points cross the 2% error, indicating that the model requires improvement. Then, through the SVM model with hyperparameter optimization, it was determined that all data points followed the diagonal line, showing the best fit of the optimized model performance.

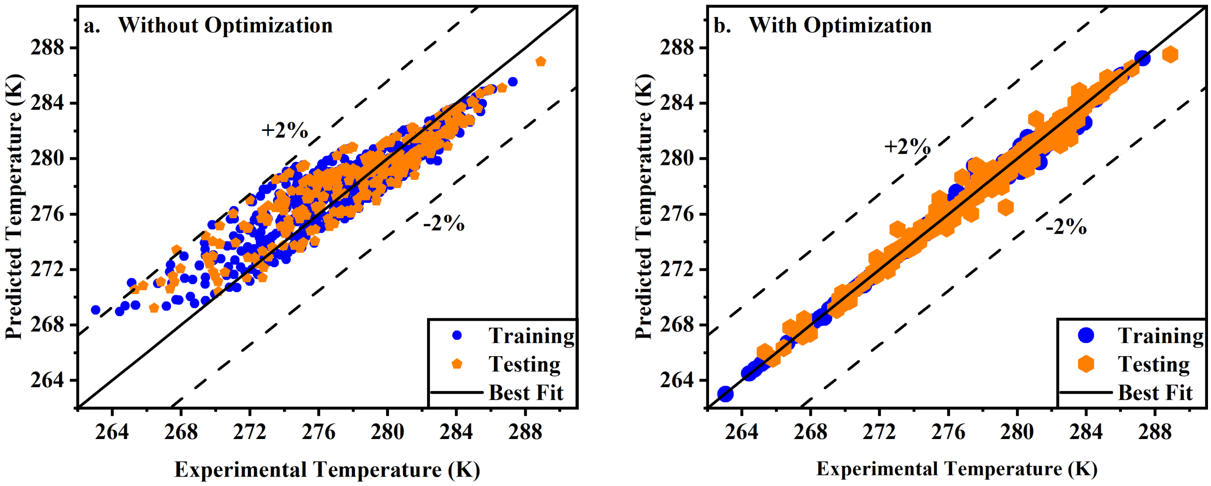

Fig. 2. Experimental versus predicted temperature of SVM model with and without optimization

3.3 Residual analysis

The residual analysis of CO_2/CH_4 hydrate inhibition using AILs with and without optimization is shown in Figure 3(a) and Figure 3(b) with and without optimization. It was found that without optimization, the residual range was -3 K to +6 K. Furthermore, a large number of training and testing data points crossed the 2 K residual, which shows that the model needs improvement. Optimization was applied, and it was found that all the training and testing data points inside the 2 K residual showed the most accurate model performance. Only one data point outside the 2 K residual can be ignored.

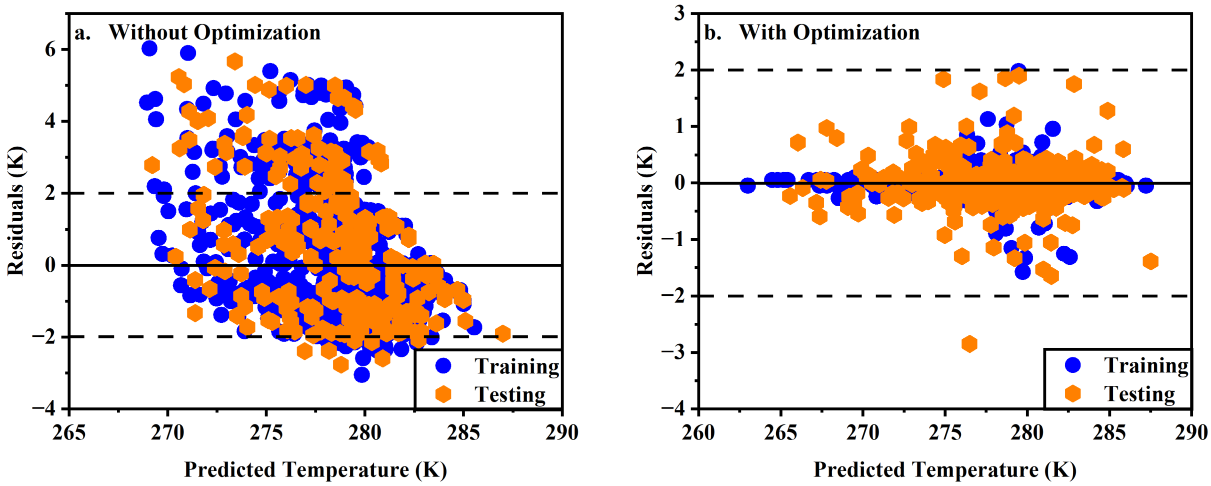

Fig. 3. Residual analysis of SVM model with and without optimization

3.4 Cumulative probability analysis

The cumulative probability versus absolute error (K) analysis for CO_2/CH_4 hydrate inhibition using ILs is shown in Figure 4(a) and Figure 4(b). It was found that without optimization, only 0.4 cumulative probability followed a 1 K absolute error, and the rest of the data went out until a 6 K absolute error. This shows that the performance of the model was poor. Furthermore, optimization was conducted on the model, and the results showed that most of the data had 0.9 cumulative probability of a 1 K absolute error. A few data points lie within a 2 K absolute error. Only two data points were outside this range. Evidently, the model performed well.

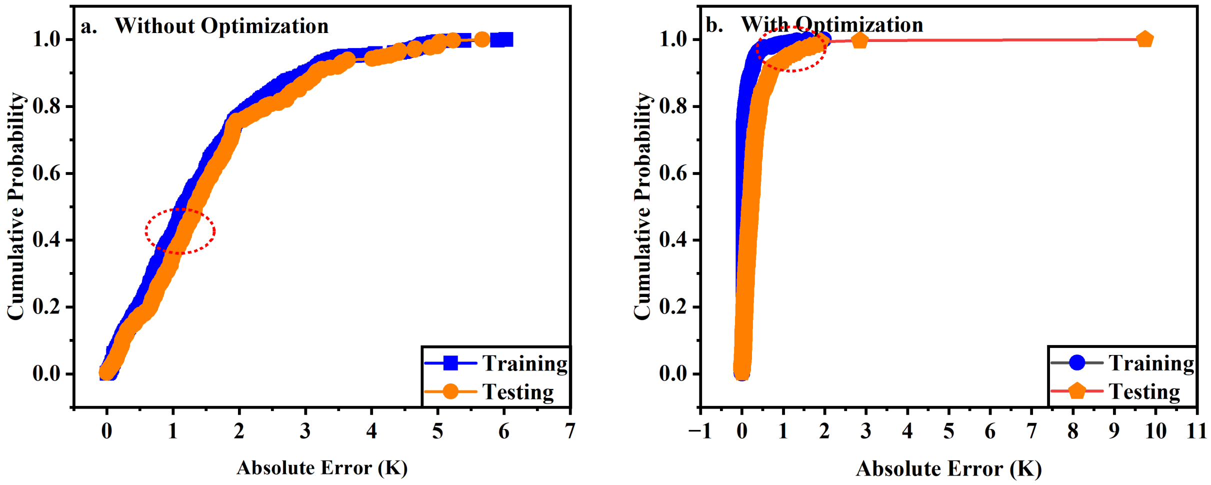

Fig. 4. Cumulative probability versus absolute temperature analysis of SVM model with and without optimization

3.5 Error Analysis

The R^2, MAE, and RMSE error analyses were performed as shown in Table 2. for CO_2/CH_4 hydrate inhibition using ILs. The R^2 analysis with optimization showed strong model performance with $R^2 > 0.97$, whereas without optimization, the results were lower, approaching 0.79. The MAE analysis for optimization approaches 0.1 and 0.3 showed strong model performance, whereas without optimization MAE is greater than 1, indicates poor performance of the model. RMSE analysis with optimization shows strong performance for the training dataset, whereas slightly poor performance for the testing dataset can still be considered good. However, without optimization, the performance was poor.

Table 2
Error Analysis of SVM model with and without optimization

	Without Optimization		**With Optimization**	
	Training	**Testing**	**Training**	**Testing**
R^2	0.7979	0.7952	0.9975	0.9710
MAE	1.4475	1.5889	0.1020	0.3411
RMSE	1.8387	1.9779	0.2062	0.7446

3.6 Model Performance for New Data

The developed SVM-optimized model for CO_2/CH_4 binary mixture hydrate inhibition using AILs was used to predict the new inhibition temperature of various CO_2/CH_4 mixtures and pure CH_4, as shown in Fig. 5.

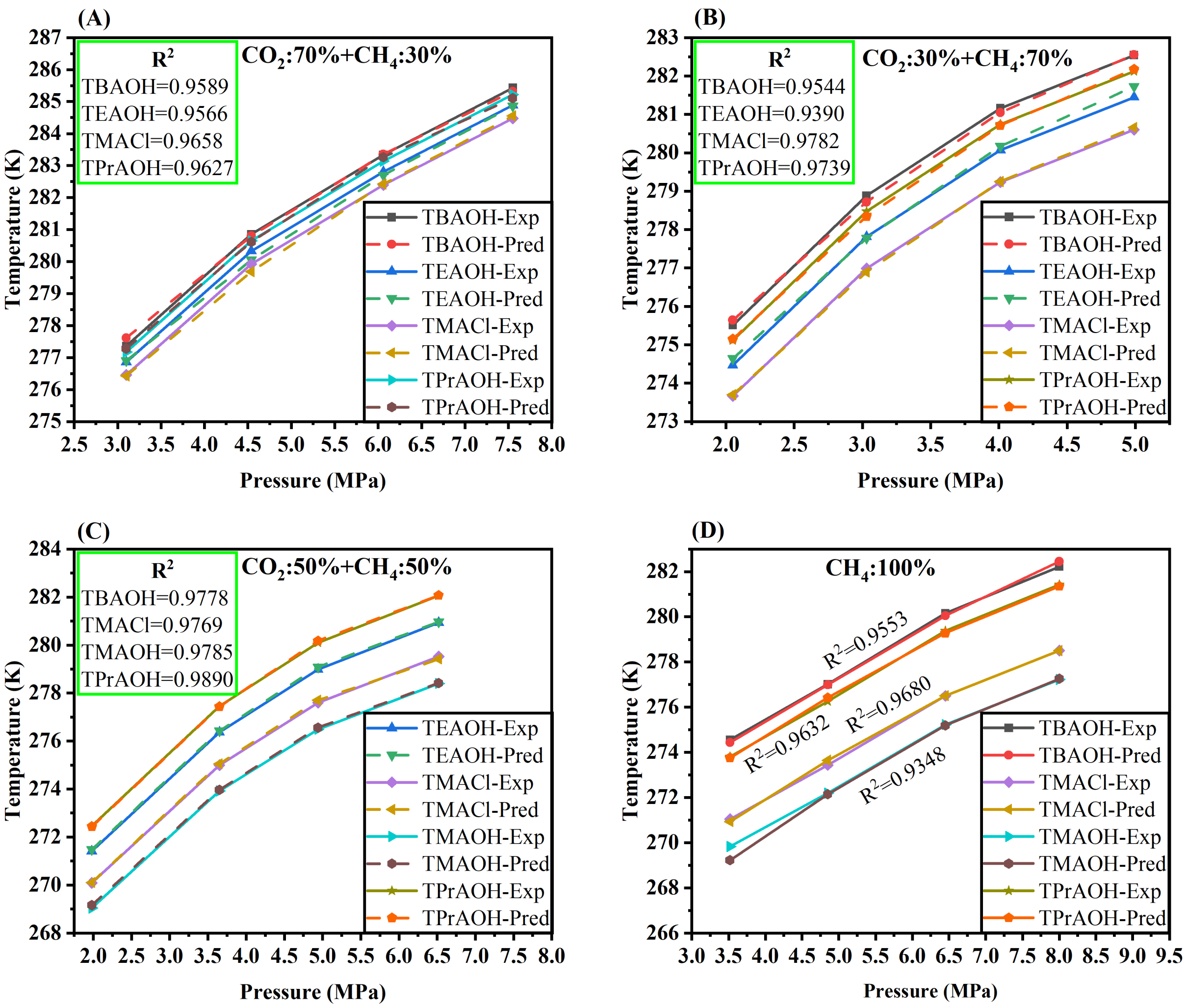

Fig. 5. The optimized SVM model performance on new data

The CO_2:70%+CH_4:30% mixture with 6 wt% of the selected inhibitors, TBAOH, TEAOH, TMACl, and TPrAOH, was used for the performance analysis, as illustrated in Fig. 5(a). The model performed well, with $R^2 > 0.95$ for all samples. Similarly, the model was applied to the CO_2:30%+CH_4:70% mixture with 12 wt% of the selected inhibitors, as described in Fig. 5(b), and it was found that the model yielded the best results with $R^2 > 0.93$. Furthermore, the same model was applied to pure CH_4 with 24 wt% of inhibitors, as in Fig. 5(d) and an equivalent mixture of CO_2/CH_4 with 21 wt% of inhibitors, as in Fig. 5(c) and it was observed that the model performance was accurate, with $R^2 > 0.93$. Hence, it can be concluded from all mixtures that the optimized SVM model performed best.

4. Conclusion

This study addresses the Support Vector Machine (SVM) modelling of CO_2/CH_4 hydrate inhibition using ammonium-based ionic liquids (AILs), which is a cost- and time-efficient alternative to experimental studies. A dataset of 981 data points was compiled, and the outliers were removed using a Williams plot. The dataset was split into training and testing datasets. The model was developed with and without hyperparameter optimization. Model performance was analyzed using experimental versus predicted temperature, residual analysis, cumulative probability versus absolute error, and error analysis metrics. Without optimization, a remarkable scatter from the diagonal line, residuals spanning from -3 to +6 K, a weak cumulative probability of 0.4 at 1 K absolute error, and lower error metrics indicated inadequate model performance. With optimization, the dataset fit the diagonal line, the residuals were limited to ±2 K, a strong cumulative probability is 0.9 at 1 K absolute error, and the lower error metrics show strong model performance. In addition, the optimized SVM model was applied to new data, and it was observed that the model performed best with $R^2 > 0.93$.

Hence, hyperparameter optimization is essential for SVM modelling of CO_2/CH_4 hydrate inhibition using AILs. In the future, the same model can be enhanced to predict the hydrate inhibition temperature for various ILs. Furthermore, the enhanced model can be used for industrial processes to inhibit hydrates using commercial inhibitors.

Acknowledgement

Technical support from the Centre for Sustainable Energy, Environment and Social Innovation (SEESI), Universiti Teknologi PETRONAS, and research funding from Yayasan University Teknologi PETRONAS (015LC0475) are highly acknowledged.

Conflict of Interest

There is no conflict of interest

Author Contribution

Dileep Kumar: Writing – original draft
Muhammad Saad Khan: Reviewing and Editing
Bhajan Lal: Supervision

Data availability

Literature and thermodynamic model data can be provided on request.

References

[1] Lal, B., and O. Nashed. *Chemical Additives for Gas Hydrates*. Springer Nature, 2019.

[2] Zulu, N. M., H. Hashemi, K. Tumba, and V. T. Adeleke. "Thermodynamic Inhibition of CO2–CH4 Gas Hydrates by DESs: Experimental and Computational Study." *Journal of Chemical & Engineering Data* 70 (2025): 3675–3689. https://doi.org/10.1021/acs.jced.5c00067.

[3] Mokhatab, S., R. J. Wilkens, and K. J. Leontaritis. "A Review of Strategies for Solving Gas-Hydrate Problems in Subsea Pipelines." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 29 (2007): 39–45. https://doi.org/10.1080/009083190933988.

[4] Elhenawy, S., M. Khraisheh, F. Almomani, M. A. Al-Ghouti, M. K. Hassan, and A. Al-Muhtaseb. "Towards Gas Hydrate-Free Pipelines: A Comprehensive Review of Gas Hydrate Inhibition Techniques." *Energies* 15 (2022): 8551. https://doi.org/10.3390/en15228551.

[5] Kumar, D., B. Lal, I. M. Saaid, K. S. Foo, S. Ridha, and N. A. Basnih. "A Simulation Study of CO2 Hydrate Inhibition in CO2+CH4 Seawater Mixture." *Engineering, Technology & Applied Science Research* 16 (2026): 31923–31929. https://doi.org/10.48084/etasr.12311.

[6] Tariq, M., D. Rooney, E. Othman, S. Aparicio, M. Atilhan, and M. Khraisheh. "Gas Hydrate Inhibition: A Review of the Role of Ionic Liquids." *Industrial & Engineering Chemistry Research* 53 (2014): 17855–17868. https://doi.org/10.1021/ie503559k.

[7] Gao, H., M. Luo, J. Xing, Y. Wu, Y. Li, W. Li, Q. Liu, and H. Liu. "Desulfurization of Fuel by Extraction with Pyridinium-Based Ionic Liquids." *Industrial & Engineering Chemistry Research* 47 (2008): 8384–8388. https://doi.org/10.1021/ie800739w.

[8] Khan, M. S., C. B. Bavoh, B. Partoon, B. Lal, M. A. Bustam, and A. M. Shariff. "Thermodynamic Effect of Ammonium Based Ionic Liquids on CO2 Hydrates Phase Boundary." *Journal of Molecular Liquids* 238 (2017): 533–539. https://doi.org/10.1016/j.molliq.2017.05.045.

[9] Rahanu, B. L., M. R. Atta, and K. S. Foo. *Artificial Intelligence and Machine Learning in Flow Assurance*. Cambridge Scholars Publishing, 2025.

[10] Lal, B., C. B. Bavoh, and J. K. S. Sayani. *Machine Learning and Flow Assurance in Oil and Gas Production*. Cham: Springer, 2023.

[11] Eslamimanesh, A., F. Gharagheizi, M. Illbeigi, A. H. Mohammadi, A. Fazlali, and D. Richon. "Phase Equilibrium Modeling of Clathrate Hydrates of Methane, Carbon Dioxide, Nitrogen, and Hydrogen+Water Soluble Organic Promoters Using Support Vector Machine Algorithm." *Fluid Phase Equilibria* 316 (2012): 34–45. https://doi.org/10.1016/j.fluid.2011.11.029.

[12] Yarveicy, H., and M. M. Ghiasi. "Modeling of Gas Hydrate Phase Equilibria: Extremely Randomized Trees and LSSVM Approaches." *Journal of Molecular Liquids* 243 (2017): 533–541. https://doi.org/10.1016/j.molliq.2017.08.053.

[13] Xu, H., Z. Jiao, Z. Zhang, M. Huffman, and Q. Wang. "Prediction of Methane Hydrate Formation Conditions in Salt Water Using Machine Learning Algorithms." *Computers & Chemical Engineering* 151 (2021): 107358. https://doi.org/10.1016/j.compchemeng.2021.107358.

[14] Khan, M. S., C. B. Bavoh, B. Partoon, O. Nashed, B. Lal, and N. B. Mellon. "Impacts of Ammonium Based Ionic Liquids Alkyl Chain on Thermodynamic Hydrate Inhibition for Carbon Dioxide Rich Binary Gas." *Journal of Molecular Liquids* 261 (2018): 283–290. https://doi.org/10.1016/j.molliq.2018.04.015.

[15] Khan, M. S., B. Lal, B. Partoon, L. K. Keong, A. B. Bustam, and N. B. Mellon. "Experimental Evaluation of a Novel Thermodynamic Inhibitor for CH4 and CO2 Hydrates." *Procedia Engineering* 148 (2016): 932–940. https://doi.org/10.1016/j.proeng.2016.06.433.

[16] Li, X.-S., Y.-J. Liu, Z.-Y. Zeng, Z.-Y. Chen, G. Li, and H.-J. Wu. "Equilibrium Hydrate Formation Conditions for the Mixtures of Methane + Ionic Liquids + Water." *Journal of Chemical & Engineering Data* 56 (2011): 119–123. https://doi.org/10.1021/je100987q.

[17] Khan, M. S. "Investigation of Ammonium Based Ionic Liquids as a Dual Functional Gas Hydrate Inhibitors." PhD diss., Chemical Engineering Department, Universiti Teknologi PETRONAS, 2019. UTPedia Repository

[18] Dickens, G. R., and M. S. Quinby-Hunt. "Methane Hydrate Stability in Pore Water: A Simple Theoretical Approach for Geophysical Applications." *Journal of Geophysical Research: Solid Earth* 102 (1997): 773–783. https://doi.org/10.1029/96JB02941.

[19] Mesbah, M., E. Soroush, and M. Rezakazemi. "Development of a Least Squares Support Vector Machine Model for Prediction of Natural Gas Hydrate Formation Temperature." *Chinese Journal of Chemical Engineering* 25 (2017): 1238–1248. https://doi.org/10.1016/j.cjche.2016.09.007.

[20] Hsu, C. Y., J. S. Buñay Guaman, A. Ved, A. Yadav, G. Ezhilarasan, A. Rameshbabu, A. Alkhayyat, D. Aulakh, S. Choudhury, S. K. Sunori, and F. Ranjbar. "Prediction of Methane Hydrate Equilibrium in Saline Water Solutions Based on Support Vector Machine and Decision Tree Techniques." *Scientific Reports* 15 (2025): 1–16. https://doi.org/10.1038/s41598-025-95969-w.

[21] Devroye, L., L. Györfi, and G. Lugosi. "Vapnik-Chervonenkis Theory." In *A Probabilistic Theory of Pattern Recognition*, 187–213, 1996. https://doi.org/10.1007/978-1-4612-0711-5_12.

[22] Liu, J., M. Bai, N. Jiang, and D. Yu. "Structural Risk Minimization of Rough Set-Based Classifier." *Soft Computing* 24 (2020): 2049–2066. https://doi.org/10.1007/s00500-019-04038-8.

[23] Behnam Motlagh, M. A., R. Hashemi, Z. Taheri Rizi, M. Mohammadi, M. Mohammadtaheri, and B. Zarei Eslam. "Comprehensive Review of the Impact of Thermodynamic Inhibitors and the Predictive Power of Machine Learning Models on Hydrate Formation Pressure and Temperature." *Journal of Chemical & Engineering Data* 70 (2025): 3891–3943. https://doi.org/10.1021/acs.jced.5c00025.

[24] Ghaddar, B., and J. Naoum-Sawaya. "High Dimensional Data Classification and Feature Selection Using Support Vector Machines." *European Journal of Operational Research* 265 (2018): 993–1004. https://doi.org/10.1016/j.ejor.2017.08.040.

[25] Yu, B., Y. T. Wang, J. B. Yao, and J. Y. Wang. "A Comparison of the Performance of ANN and SVM for the Prediction of Traffic Accident Duration." *Neural Network World* 26 (2016): 271–287. https://doi.org/10.14311/NNW.2016.26.015.