ISSN 2411-3336; e-ISSN 2541-9404

JOURNAL OF MINING INSTITUTE
Zapiski Gornogo instituta
Journal homepage: pmi.spmi.ru

Research article
Geotechnical Engineering and Engineering Geology

Predicting gas hydrate formation temperature using a hybrid artificial intelligence-empirical approach

Mohsen Riazi[1], **Hossein Mehrjoo**[1], **Mohammad-Saber Dabiri**[1], **Jamilyam Ismailova**[2], **Masoud Riazi**[3]✉
[1] *Shahid Bahonar University of Kerman, Kerman, Iran*
[2] *Satbayev University, Almaty, Republic of Kazakhstan*
[3] *Nazarbayev University, Astana, Republic of Kazakhstan*

***How to cite this article*:** Riazi M., Mehrjoo H., Dabiri M.-S., Ismailova J., Riazi M. Predicting gas hydrate formation temperature using a hybrid artificial intelligence-empirical approach. Journal of Mining Institute. 2026. Vol. 280, p. 145-162.

Abstract
Accurate prediction of gas hydrate formation temperature is essential for ensuring flow assurance during natural gas production and subsea transportation. In this study, 459 experimental data points collected from the literature, covering a wide pressure range of 330-68,600 kPa and gas specific gravity range of 0.552-1.03, were utilized to develop reliable predictive tools using only two practical input variables: pressure and gas specific gravity. Two complementary modeling approaches were implemented. First, a systematic comparative analysis of several advanced machine learning technique, including Multilayer Perceptron (MLP) trained with Levenberg – Marquardt (LM), Bayesian Regularization (BR), and Scaled Conjugate Gradient (SCG) algorithms; Radial Basis Function (RBF); CatBoost; XGBoost; and Least-Squares Support Vector Machine (LSSVM), was conducted. Second, an optimized empirical correlation was formulated, with its coefficients determined using the Generalized Reduced Gradient (GRG) nonlinear optimization algorithm to minimize overall prediction error across the entire experimental dataset. Among the machine learning models, CatBoost demonstrated superior predictive performance, achieving an AAPRE of 0.094 % and an R^2 value of 0.996 on the testing dataset. The developed GRG-based correlation also exhibited strong agreement with experimental data while providing a transparent and computationally efficient alternative suitable for rapid engineering calculations. Sensitivity and leverage analyses confirmed the robustness, statistical reliability, and broad applicability of the proposed models within the investigate domain.

Keywords
gas hydrate temperature; least-squares support vector machine; machine learning; multilayer perceptron; radial basis function; generalized reduced gradient

Funding
None.

Received: 09.02.2025 ***Accepted: 25.06.2026*** ***Online: 28.08.2026*** ***Published: 10.09.2026***

Introduction

In recent years, natural gas has emerged as an abundant, valuable, and environmentally friendly energy source [1-4]. Substantial efforts are being made to promote and expand its utilization across various sectors, driven not only by its technical and economic advantages but also by its low greenhouse gas emissions [2, 5]. Nevertheless, natural gas production often occurs under unconventional conditions, presenting numerous operational challenges. Extraction, together with associated field operations such as deep-water drilling, pipeline transportation, storage, and separation, faces significant flow assurance issues [6-9]. Among these, hydrate formation is recognized as the most critical, as it can cause severe consequences, including blockages in pipelines and blowout preventers, as well as operational instability and potential safety and integrity failures in field

This is an open access article under the CC BY 4.0 license

installations [10-13]. Gas hydrates are crystalline solids with a clathrate structure, formed when small gas molecules, referred to as "guests," including hydrogen sulfide, carbon dioxide, methane, ethane, propane, and nitrogen, become encaged within cavities created by water molecules, known as the "host." These hydrates generally form under high-pressure and low-temperature conditions [14-19]. The formation of hydrates is frequently observed due to the continuous presence of water during gas production or as a result of improper pipeline design and construction.

In multiphase transfer pipelines, surface facilities, and wellheads, hydrates present a major flow assurance challenge. The effective and timely implementation of hydrate prevention strategies is critical, as blockages caused by hydrate formation can halt production for days or even months. Such disruptions may result in substantial economic losses, potentially amounting to millions of dollars in gas gathering systems or high-pressure natural gas transmission pipelines. E.Hammerschmidt et al. [20] first reported hydrate plugging in pipelines in 1934. Since then, the deposition of hydrates during the transportation and production of natural gas has become a prominent area of investigation. Considerable advancements in hydrate research have drawn significant attention, emphasizing the importance of understanding hydrate formation, identifying its key parameters, and developing effective mitigation strategies for both academic research and industrial practice [17, 21-26]. Given the multidisciplinary relevance of hydrate formation, there has been increasing interest in accurately determining its formation conditions [27]. Among these, the hydrate formation temperature is one of the most critical parameters [28, 29]. While experimental methods provide reliable predictions of hydrate formation temperature, they are often costly and time-consuming. As a result, researchers have pursued more cost-effective predictive approaches, including empirical correlations, chart-based techniques, and theoretical models based on various equations of state [27-29].

Although the earliest model for predicting hydrate formation temperature, the correlation introduced by E.Hammerschmidt in 1934 [20], elated hydrate formation temperature solely to pressure, subsequent and more advanced correlations have incorporated both pressure and gas composition as key input parameters [30-35]. Building on these developments, several researchers have also employed machine learning techniques to predict gas hydrate formation temperature, offering enhanced predictive accuracy and efficiency compared to traditional methods.

A.A.Elgibaly and A.M.Elkamel [36] developed a sophisticated artificial neural network (ANN) model to predict hydrate formation conditions for various pure gases, gas mixtures, and different inhibitors. Their study examined multiple neural network architectures to identify the most accurate configuration. According to their findings, a network with a single hidden layer containing 25 neurons provided reasonably precise results for the gravity-based model. For compositional models, networks with 40 or 50 neurons in the hidden layer yielded the lowest prediction errors. In addition, they introduced a novel ANN model based on gas gravity, incorporating updated experimental data on hydrate formation conditions. This model demonstrated superior accuracy compared with conventional gravity-graphic methods and empirical correlations. However, for gas gravities between 0.8 and 1.0, its applicability was limited due to the scarcity of experimental data. For gravities exceeding 1.0, the model produced ambiguous results, likely as a consequence of insufficient data and the limitations of relying solely on gas gravity as an input parameter.

G.Zahedi et al. [37] employed an ANN-based approach to predict hydrate formation temperature and reported outstanding predictive performance. Their results showed that the ANN model achieved higher accuracy compared with conventional prediction methods.

This is an open access article under the CC BY 4.0 license

S.H.Riazi et al. [38], using experimental data from E.D.Sloan and C.A.Koh [15], developed a new correlation through MATLAB curve-fitting techniques. The correlation was further optimized using the Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and the Imperialist Competitive Algorithm (ICA) to enhance its accuracy. Compared with previous models, the proposed correlation exhibited the lowest error and the highest overall accuracy.

S.M.Hesami et al. [39] introduced two advanced ANN-based techniques optimized using ICA to predict gas hydrate formation conditions. The ICA-ANN model was developed using empirical data from the literature and evaluated against a conventional ANN model as well as a rigorous thermodynamic model under various operating conditions. Unexpectedly, the ICA-ANN model underperformed relative to the conventional ANN approach.

M.N.Amar [40] proposed a reliable and user-friendly correlation to estimate hydrate formation temperature for various types of natural gas, including sour, acidic, and sweet gases. This correlation was developed using Gene Expression Programming (GEP), an explicit machine learning technique. The results indicated that the GEP-based model predicted hydrate formation temperature with high accuracy, achieving an average absolute relative error (AAPRE) of only 0.1397 %.

M.Mesbah et al. [41] developed a universal and user-friendly correlation based on a hybrid Group Method of Data Handling (GMDH) approach to predict hydrate formation temperature for a wide range of natural gas mixtures, including both sweet and sour gases. Compared with traditional correlations, this model demonstrated superior statistical performance. Notably, it showed promising results for sour gases with high H_2S concentrations, outperforming existing thermodynamic models and empirical correlations. A leverage statistical method was applied to identify questionable data points and assess model reliability. The analysis confirmed that nearly all data points fell within the valid applicability range, supporting the statistical robustness of the proposed hybrid GMDH model.

To predict methane hydrate formation temperature over a broad range of brine concentrations, M.Hosseini and Yu.Leonenko [42] proposed machine learning-based models. Their findings indicated that extremely randomized trees achieved strong predictive performance, with a testing root mean square error (RMSE) of 0.6248.

In 2024, M.Karaköse and Ö.Yücel [43] developed advanced machine learning regression models to estimate hydrate formation temperature based on the chemical composition of sweet gas mixtures. The applied techniques included linear regression, decision tree regression (DT), random forest regression (RF), generalized additive models, and ANN. Ensemble methods, particularly random forest, demonstrated superior performance, achieving an R^2 and adjusted R^2 of 0.998. These results highlight the potential of ensemble learning approaches to improve the safety and efficiency of gas transportation and processing systems.

N.Khan et al. [44] evaluated three machine learning algorithms – Support Vector Regression (SVR), ANN, and DT – for predicting hydrate formation temperature. Their results showed R^2/RMSE values of (0.9999; 0.0631), (0.9986; 0.5011), and (0.9278; 3.5606) for SVR, ANN, and DT, respectively. Based on these findings, SVR exhibited the highest predictive accuracy, followed by ANN, while DT showed comparatively lower performance.

W.Aleem et al. [45] presented a comprehensive and novel comparative analysis of nine distinct machine learning (ML) models for the accurate prediction of methane hydrate formation temperatures in the presence of brines. The investigations applied major ML algorithms, including Multiple Linear Regression (MLR), Long Short-Term Memory (LSTM), Radial Basis Function (RBF), Support Vector Machine (SVM), ANN, Gradient Possess Regression (GPR), Random Forest (RF) and K-Nearest Neighbors (KNN). Model accuracy was validated against a large dataset comprising 1000 data points

This is an open access article under the CC BY 4.0 license

with a diverse range of salt concentrations. The experimental results exhibited the KNN algorithm as fast-converging, accurate, and consistent across the entire range of data points, achieving and R^2 score of 0.975 and an AAPRE of 0.385 %.

M.Khafaie et al. [46] in their study developed an evaluation of two ML models including ANN and hybrid Particle Swarm Optimization-ANN (PSO-ANN). The approach centered on predicting hydrate formation temperature using key input parameters, namely the mole fraction of gas components and system pressure. A dataset comparing 1841 experimental samples was utilized. The ANN model achieved a Mean Squared Error (MSE) of 0.0003533 and R^2 of 0.9976. The PSO-ANN model yielded an MSE of 0.0003298 and R^2 of 0.9899. Both models demonstrated high predictive accuracy, with the PSO-ANN exhibiting a slight advantage in MSE. This research highlighted the efficacy of a streamlined, optimized ML approach for accurate and generalizable gas hydrate temperature.

M.R.Atta et al. [47] conducted a systematic review following PRISMA guidelines and bibliometric mapping to analyze studies on Artificial Intelligence and ML for predicting gas hydrate thermodynamic conditions published between 2016 and 2025. The analysis indicated a predominant focus on binary water – gas system (87.5 %) and a strong reliance on experimental literature-derived datasets (73.9 %). ANN, SVR and RF models were the most prevalent algorithms, with R^2 and RMSE as primary evaluation metrics. A significant limitation identified was the lack of uncertainly quantification and the under-representation of field-derived datasets. The review highlighted opportunities for future work in evaluating model robustness expanding model outputs to include hydrate kinetics and morphology. Although extensive experimental and numerical investigations have been conducted on hydrate formation temperature, there remains a need for comprehensive studies that simultaneously develop accurate, robust, and practically applicable machine learning models and simplified correlations.

In this study, a comprehensive dataset comprising 459 data points was used to develop advanced predictive models for estimating gas hydrate formation temperature. In the first approach, several robust ML techniques – including CatBoost, XGBoost, Least-Squares Support Vector Machine (LSSVM), RBF network, and Multilayer Perceptron (MLP) – were implemented. The models were trained using the Levenberg – Marquardt (LM), Bayesian Regularization (BR), and Scaled Conjugate Gradient (SCG) algorithms. In the second approach, an empirical correlation was developed using the Generalized Reduced Gradient (GRG) optimization technique, accounting for variations in pressure and gas specific gravity. The objective was to derive a novel correlation capable of achieving superior accuracy and performance compared with existing models. A relevancy factor analysis was conducted to assess the influence of input parameters on gas hydrate formation temperature. Furthermore, the leverage method was applied for statistical analysis to identify potential outliers and determine the domain of applicability of the developed models.

Data

The generalizability and reliability of the developed models are strongly influenced by the diversity and quantity of the input data used during model development. In this study, a comprehensive databank comprising of 459 experimental data points, covering a wide range of operating conditions, was compiled to develop accurate predictive models for gas hydrate formation temperature [48-55]. presents a statistical summary of the input variables, including the gas specific gravity and pressure, as well as the output variable (the experimental temperature). The complete dataset used in this study is provided in Supplement https://pmi.spmi.ru/pmi/article/download/16663/16705.

This is an open access article under the CC BY 4.0 license

Table 1

Statistical description of the employed database [48-56]

Features	Gas specific gravity	Pressure, kPa	Experimental temperature, K
Maximum	1.03	68600	304.483
Minimum	0.552	330	272.04
Skewness	1.555	3.474	0.265
Kurtosis	2.338	14.396	–0.809
Mode	0.629	20684.272	283.15
Median	0.651	3164.999	284.349
Mean	0.685	6722.435	284.956
Standard deviation	0.097	10136.045	7.52

As shown in Table 1, the gas specific gravity ranges from 0.552 to 1.03. The pressure varies between 330 and 68600 kPa, with an average value of 6722.435 kPa. The experimental temperatures in the databank range from 272.04 to 304.483 K, with an average of 284.956 K.

Methods

In this study, several machine learning methods were employed to predict gas hydrate formation temperature.

Least-Squares Support Vector Machine (LSSVM). The LSSVM algorithm is widely used for regression analysis, pattern recognition, and data classification tasks. LSSVM retains the main characteristics of the traditional SVM while offering a simplified formulation that reduces computational complexity and training time. Detailed information on SVM theory can be found in the literature [57]. The LSSVM model was first introduced by J.A.K.Suykens and J.Vandewalle [58, 59]. The cost function of this method is similar to that of SVM and expressed as follows:

$$g(y) = O^T \varphi(y_i) + bi,$$

where bi – biased terms; O^T – transposed output layer vector; $\varphi(y_i)$ – kernel function; y_i – input vector of the function.

The input matrix has dimensions $M \times m$, where m and M represent the number of input parameters and the total number of data points, respectively. By minimizing the cost function, the bias term and the output layer vector can be determined [60]:

$$\mathrm{CF} = \frac{1}{2} O^T O + \frac{1}{2} \gamma \sum_{i=1}^{num} s_i^2.$$

The cost function [60]:

$$x_i = O^T \varphi(y_i) + bi + s_i^2, \quad i = 1, 2, 3, ..., M,$$

where s_i – regression error; γ – regularization (adjustable) parameter in LSSVM.

The Lagrangian form of the problem [60]:

$$K(O, s, q, bi) = \frac{1}{2} O^T O + \frac{1}{2} \gamma \sum_{i=1}^{num} s_i^2 - \sum_{i=1}^{num} q_i \left(O^T \varphi(y_i) + bi + s_i - x_k \right),$$

where q_i – Lagrange multiplier.

By taking the partial derivatives of the Lagrangian with respect to bi, q_i, s_i, and O, the unknown parameters can be obtained [60]:

This is an open access article under the CC BY 4.0 license

© Mohsen Riazi, Hossein Mehrjoo, Mohammad-Saber Dabiri, Jamilyam Ismailova, Masoud Riazi, 2026

$$\begin{cases} \dfrac{\partial K}{\partial O} = 0 \to O = \sum\limits_{i=1}^{num} q_i \varphi(y_i); \\ \dfrac{\partial K}{\partial bi} = 0 \to \sum\limits_{i=1}^{num} q_i = 0; \\ \dfrac{\partial K}{\partial s_i} = 0 \to q_i = \gamma s_i, \ \ i = 1,2,3,...,M; \\ \dfrac{\partial K}{\partial q_i} = 0 \to O^T \varphi(y_i) + bi + s_i - x_k = 0, \ \ i = 1,2,3,...,M. \end{cases}$$

Solving this system of equations yields the required variables in the LSSVM framework. The following equations present three commonly used kernel functions:

- radial basis function kernel – $\varphi(y, y_i) = \exp\left(\dfrac{-\|y_i - y\|^2}{\sigma^2}\right)$;
- linear kernel – $\varphi(y, y_i) = y_i^T y$;
- polynomal kernel – $\varphi(y, y_i) = \left(\tau + y_i^T y\right)^n$.

In LSSVM, σ^2 is an adjustable parameter that can be optimized by minimizing an objective function, such as the mean squared error. Consequently, by using the following equations, the difference between the predicted outputs of the developed model and the corresponding experimental values is minimized as follows:

$$\text{MSE} = \frac{\sum\limits_{j=1}^{m}\left(R_{exp_j} - R_{\text{rep/pred}_j}\right)^2}{m}.$$

Radial Basis Function (RBF). The RBF network is a well-established type of feedforward artificial neural network and is widely used as a robust modeling approach for solving complex problems [61, 62]. An RBF neural network typically consists of three layers: an input layer, a hidden layer, and an output layer. In the hidden layer, each neuron is associated with a center and a predefined radius, defining a specific region in the input space. Within each neuron, the distance between the input vector and the corresponding center is calculated [63]. For this purpose, various activation functions, such as thin-plate spline, Gaussian, cubic, logistic, and quadratic functions, are commonly employed [64]. The Euclidean distance between the input vector and the center of each neuron is calculated using [65]

$$r_i = \left[\sum_{j=1}^{l}\left(y_i - Ce_{ji}\right)^2\right]^{0.5}, \ \ i = 1, ..., M,$$

where the subscripts i and l – number of center vectors and the number of inputs, respectively; y_i and r_i – input vector and the Euclidean distance from the neuron's center; Ce_{ji} – cluster center.

The output of the RBF model [66] as follows, further details regarding this neural network model can be found in references [67-69]:

$$x_c(y) = \sum_{i=1}^{M} cw_i \varphi_{ci}\left(\|y_c - Ce_i\|\right), \quad i = 1, ..., M, \ \ c = 1, ..., N,$$

where cw_i – connection weight; M – number of hidden-layer neurons; Ce_i – center vector; φ – transfer function.

This is an open access article under the CC BY 4.0 license

Multilayer Perceptron (MLP). The MLP is a widely used feedforward ANN [65] composed of an input layer, one or more hidden layers, and an output layer. The input variables are fed into the input layer, while the hidden layers, typically employing nonlinear activation functions such as tansig and logsig, transform these inputs into higher-level representations [70, 71]. The output layer generally utilizes a purelin activation function to generate the final predictions. Training the MLP is performed using the backpropagation algorithm, in which the network weights and biases are iteratively updated to minimize the error between the predicted and target outputs. Although a single hidden layer is often sufficient to capture the relationships between inputs and outputs, additional layers may be required when modeling more complex systems.

Generalized Reduced Gradient (GRG). The GRG method is widely used to solve optimization problems involving multiple variables. This approach is based on the concept of reduced gradients and is capable of handling both linear and nonlinear constraints while ensuring that active constraints are satisfied throughout the solution process. By approximating the gradient at a given point z, the GRG method simultaneously considers the gradients of both the objective function and the constraint equation. As a result, the constraints are incorporated into the gradients of the objective function, restricting the search to feasible directions and effectively reducing the solution space. In this study, $f(z)$ represents the objective function to the constraint $h(z)$ [72].

Minimizes: $f(z) = z$; subjected to: $h_k(z) = 0$. The GRG formulation can be expressed as follows:

$$\frac{df}{dz_k} = \nabla z_k^t f - \nabla z_i^t f \left(\frac{dh}{dz_i} \right) \frac{dh}{dz_k}.$$

A necessary condition for minimizing $f(y)$ is that its first derivative vanishes, i.e., $df(y) = 0$.

Optimization algorithm. The effectiveness of neural network models largely depends on the selection of appropriate training algorithms used to optimize their weights and biases. In this study, three optimization algorithms were employed to train the MLP model. The LM algorithm is widely recognized for its capability to enhance MLP performance by efficiently solving nonlinear least-squares problems [73, 74]. Furthermore, the BR algorithm provides a robust alternative by adjusting the network weights and biases within the LM training framework, while applying Bayes' theorem to update their distributions [75, 76]. The SCG algorithm eliminated the need for computationally expensive line searches, thereby offering a faster and more efficient training process compared with other gradient-based methods.

Results and discussion

In this study, the formation temperature of gas hydrates was modeled over a wide range of operating conditions. The intelligent models used pressure and gas specific gravity as input variables, while the gas hydrates formation temperature served as the output. A comprehensive databank containing 459 experimental data points was randomly divided into two subsets for model development:

• Training Set. This subset was used for model development, optimization, and performance improvement. During training, the model establishes a relationship between input parameters and the output by iteratively adjusting biases and weights to minimize the deviation between the experimental and predicted values.

• Test Set. This subset was reserved to evaluate the predictive capabilities of model and assess their accuracy.

For all models in this research, 80 % of the data were allocated to the training phase, while the remaining 20 % were used for testing. It is important to note that good performance during training

This is an open access article under the CC BY 4.0 license

does not necessarily guarantee the model reliability. Therefore, testing on the reserved data is essential to validate the accuracy and robustness of the developed models.

Advanced machine learning algorithms, including CatBoost, XGBoost, MLP, RBF, and LSSVM, were employed to model the gas hydrate formation temperature. Additionally, three optimization techniques, LM, SCG, and BR, were applied to train the MLP network. Various statistical and graphical error analyses were conducted to evaluate the performance of the developed models and sensitivity analysis was performed to determine the influence of the input parameters on the models' outputs.

Prior to model development, all input and output variables were normalized using min-max scaling to the range of −1 to 1 to improve numerical stability and convergence. The predicted outputs were subsequently transformed back to their original scale. As an initial step, a correlation based on the input parameters was developed to estimate the formation temperature of gas hydrates.

Development of correlation. For the GRG algorithm, a correlation was established, demonstrating its capability to predict the target parameters with high accuracy. As a result, a clear equation was developed, comprising two input parameters and five additional coefficients:

$$T = aSG^{b}\left(\frac{\log(P)}{1.5}\right)^{C} - dSG\sqrt{\frac{\log(P)}{3.25}} + e, \tag{1}$$

where T – temperature, K; P – pressure, kPa; SG – gas specific gravity; a, b, c, d, and e – correlation constants, a = 162.4292; b = 0.244852; c = 0.36583; d = 50.60196; e = 116.777.

The correlation presented in equation (1) can be used to estimate the gas hydrate formation temperature as a function of pressure and gas specific gravity.

Evaluation of models. In this study, the proposed models were evaluated using statistical error analysis. Their performance was assessed using several metrics, including the coefficient of determination R^2, RMSE, and AAPRE:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(T_{exp,i} - T_{\mathrm{pred},i}\right)^2}{N}};$$

$$\mathrm{AAPRE} = \frac{100}{N}\sum_{i=1}^{N}\frac{\left|T_{exp,i} - T_{\mathrm{pred},i}\right|}{T_{exp,i}};$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{N}\left(T_{exp,i} - T_{\mathrm{pred},i}\right)^2}{\sum_{i=1}^{N}\left(T_{exp,i} - \bar{T}\right)^2},$$

where N – total number of data points; $T_{exp,i}$ – experimentally observed temperature; $T_{\mathrm{pred},i}$ – predicted temperature.

Statistical evaluation. To identify the most accurate model, the developed models were systematically evaluated and compared using statistical metrics, including R^2, AAPRE, and RMSE. According to statistical principles, model accuracy and validity improve as the R^2 value approaches 1, while lower values of RMSE and AAPRE indicate greater precision in the modeling process. The parameter values are presented in Table 2.

The results revealed that, among the tested models, the CatBoost model exhibited the highest accuracy, followed by XGBoost, MLP-LM, MLP-BR, MLP-SCG, and LSSVM in terms of predictive performance. As shown in Table 2, the AAPRE values of 0.126 % for the testing set, 0.086 % for the

This is an open access article under the CC BY 4.0 license

training set, and 0.094 % for the entire dataset indicate that the CatBoost approach provides highly accurate predictions of gas hydrate formation temperature. Additionally, this method achieved an outstanding overall R^2 value of 0.996. The results further established the following hierarchy of predictive accuracy for gas hydrate formation temperature: CatBoost outperformed XGBoost, followed by MLP-LM, MLP-BR, MLP-SCG, RBF, and finally LSSVM. Overfitting was assessed by comparing training and testing performance; the minimal difference in R^2 and AAPRE demonstrates that the developed models generalize well across the dataset.

Table 2

Statistical indexes of the developed model for the experimental temperature

Errors	MLP			RBF	LSSVM	CatBoost	XGBoost	Developed correlation
	LM	BR	SCG					
Training								
AAPRE, %	0.187	0.21	0.271	0.308	0.311	0.086	0.103	0.35
RMSE	0.753	0.86	1.037	1.12	1.14	0.414	0.405	1.44
R^2	0.99	0.987	0.981	0.978	0.977	0.997	0.997	0.96
Testing								
AAPRE, %	0.18	0.208	0.277	0.309	0.31	0.126	0.249	0.35
RMSE	0.756	0.837	1.102	1.155	1.13	0.494	1.198	1.44
R^2	0.99	0.987	0.978	0.975	0.978	0.996	0.974	0.96
Total								
AAPRE, %	0.186	0.209	0.272	0.308	0.311	0.094	0.132	0.35
RMSE	0.753	0.855	1.05	1.127	1.138	0.431	0.647	1.43
R^2	0.989	0.987	0.98	0.977	0.977	0.996	0.992	0.96

The hyperparameters of all machine learning models employed in this study were carefully selected to ensure optimal performance:

• MLP: hidden layers – [5.5]; transfer function – tansig, logsig, purelin; training function – LM, BR, SCG;

• RBF: spread – 2.78; max neuron – 100;

• LSSVM: σ^2 – 1.3130; kernel type – RBF kernel; γ – 589.5852;

• CatBoost: colsample bytree – 0.9; gamma – 0.01; learning rate – 0.05; max depth – 5; min child weight – 5; n_estimator – 1000; reg_lambda – 10; subsample – 0.9;

• XGBoost: learning rate – 0.05; l2 leaf reg – 5; iterations – 1500; depth – 6.

In addition to the ML models developed in this study, a correlation was proposed for estimating the formation temperature of gas hydrates. The performance of the proposed correlation, compared with existing models, is presented in Table 3.

Table 3

Statistical analysis of the suggested correlation relative to existing models

Model	AAPRE, %	RMSE	R^2
New correlation	0.3536	1.4287	0.9645
B.Towler, S.Mokhatab [33]	71.4943	203.552	0.9565
M.Safamirzaei et al. [35]	0.4071	1.5506	0.9607
E.Hammerschmidt et al. [20]	66.6774	190.9806	0.8660
S.Chavoshi et al. [77]	0.5292	1.9424	0.9455
M.Motiee et al. [31]	74.2416	211.3726	0.9546

This is an open access article under the CC BY 4.0 license

As shown in Table 3, the correlation developed in the present study demonstrated superior performance compared with existing correlations: AAPRE – 0.3536 %, RMSE – 1.4287, and R^2 – 0.9645.

In addition to the statistical evaluation, a visual assessment was conducted through graphical analysis to examine model performance. Figure 1 presents cross-plots of the predicted versus experimental gas hydrate formation temperatures for both the training and testing datasets, facilitating a visual assessment of the models' accuracy.

A cross-plot is a type of chart that illustrates the relationship between predicted and experimental values. As shown in Fig.1, the tight clustering of predicted data points along the unity line indicates a strong correlation between observed and predicted values, highlighting the remarkable precision of the CatBoost models. Furthermore, the greatest data dispersion is evident in the cross-plot of the LSSVM model, which predicts the gas hydrate formation temperature to be lower than that of the other developed models. This observation is consistent with the error values reported in Table 2.

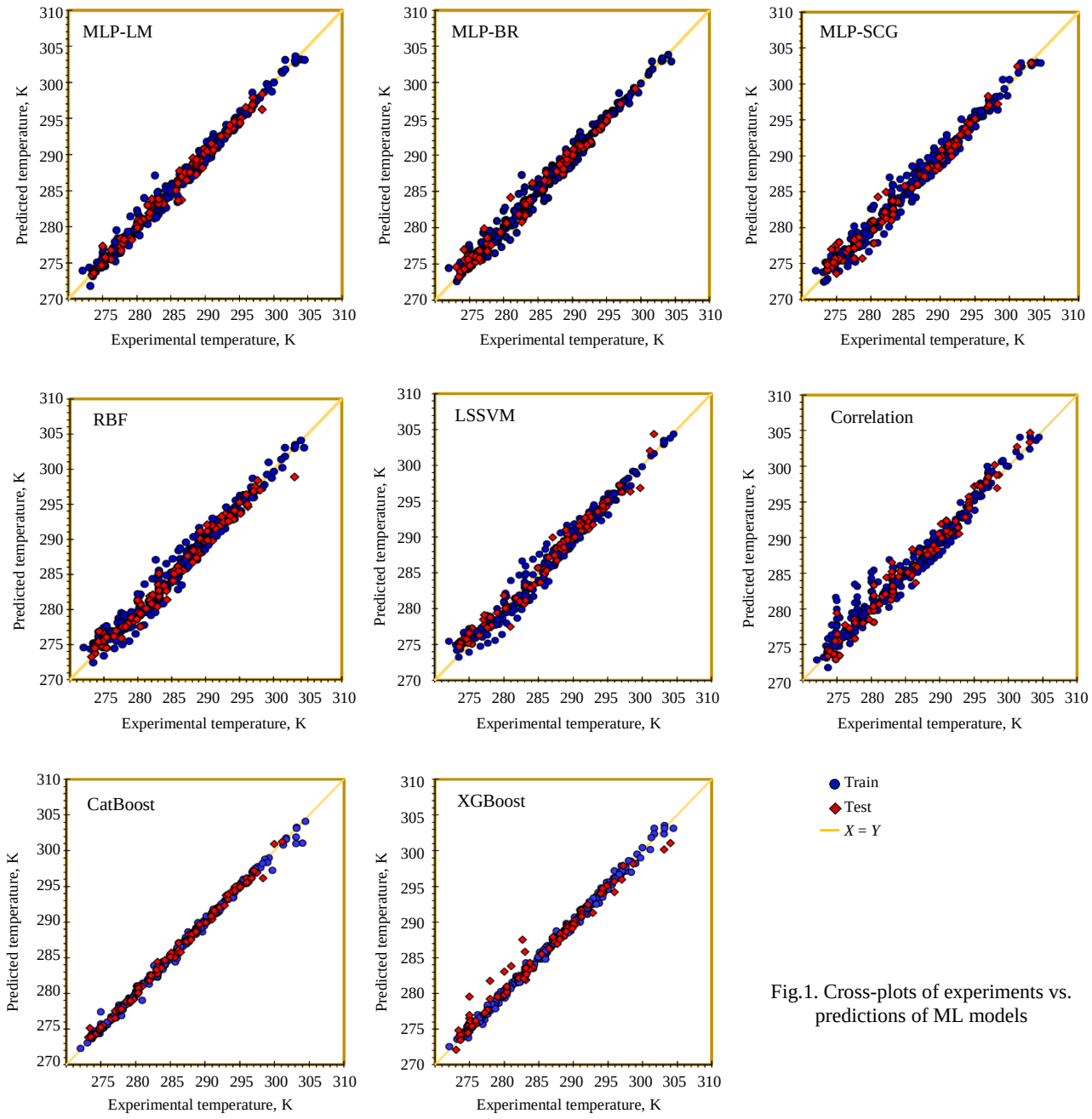


Fig.1. Cross-plots of experiments vs. predictions of ML models

This is an open access article under the CC BY 4.0 license

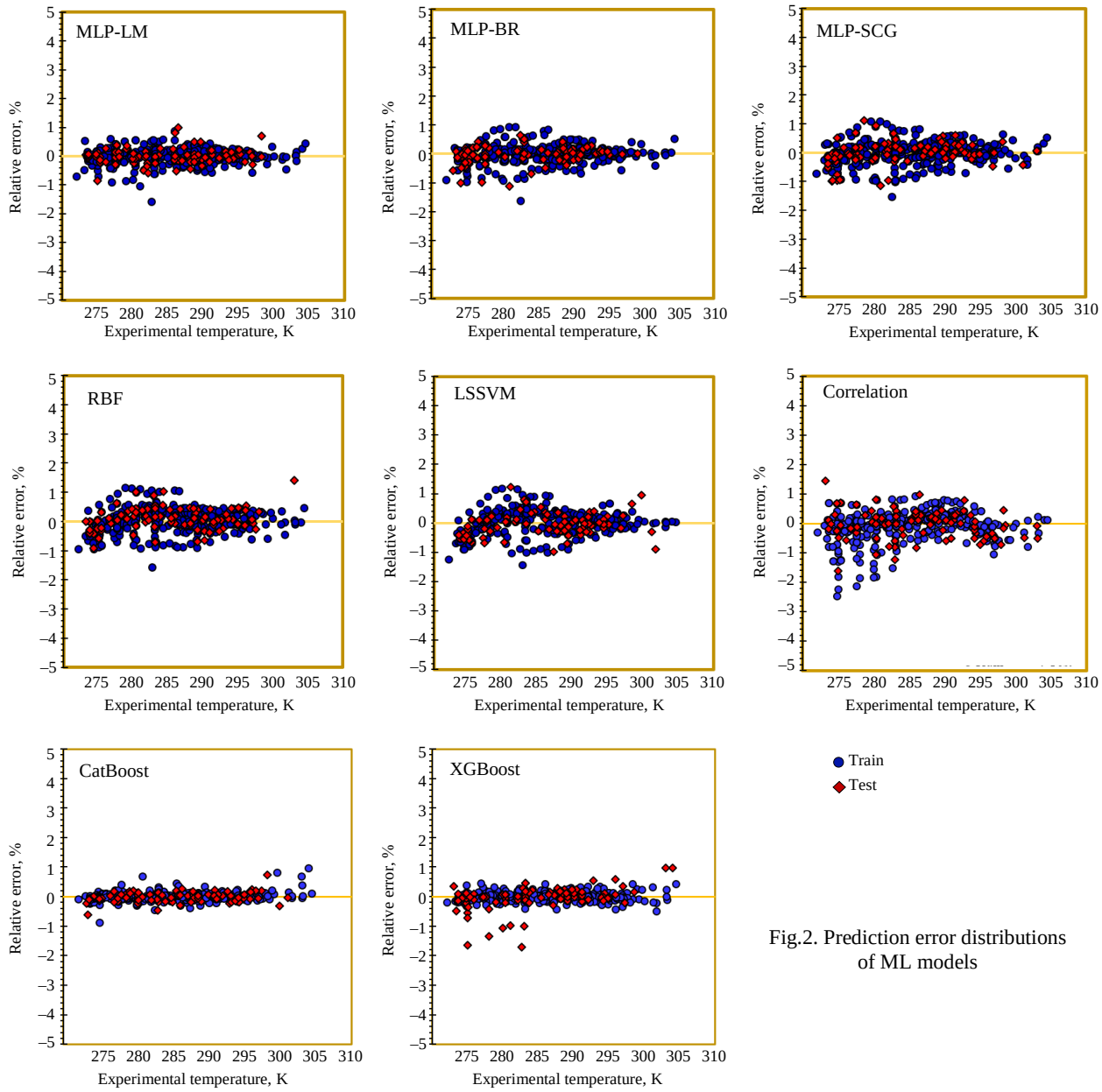


Fig.2. Prediction error distributions of ML models

Figure 2 illustrates the distribution of relative error percentages against the experimentally measured gas hydrate formation temperatures for all models. In this error evaluation, the observed output data are compared with the AAPRE values.

A predictive model demonstrates superior performance in estimating gas hydrate temperature when most data points exhibit errors close to zero. The near-zero errors of the CatBoost model confirmed its precision and reliability. However, other models, particularly XGBoost, can also provide accurate predictions with minimal errors for gas hydrate temperature estimation.

The performance of the developed models, evaluated using R^2 metrics, is shown in Fig.3.

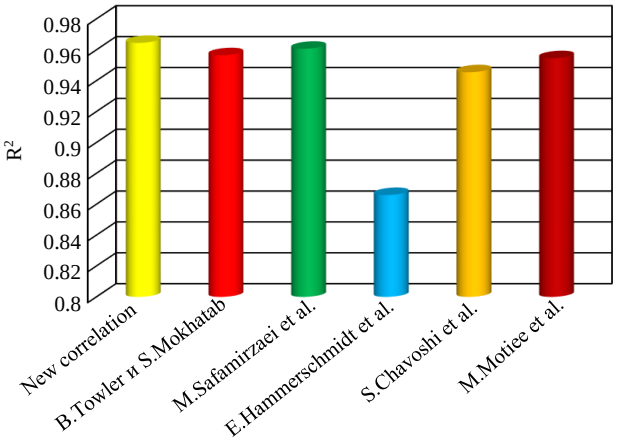


Fig.3. Comparison of R^2 values among the suggested correlation models

This is an open access article under the CC BY 4.0 license

© Mohsen Riazi, Hossein Mehrjoo, Mohammad-Saber Dabiri, Jamilyam Ismailova, Masoud Riazi, 2026

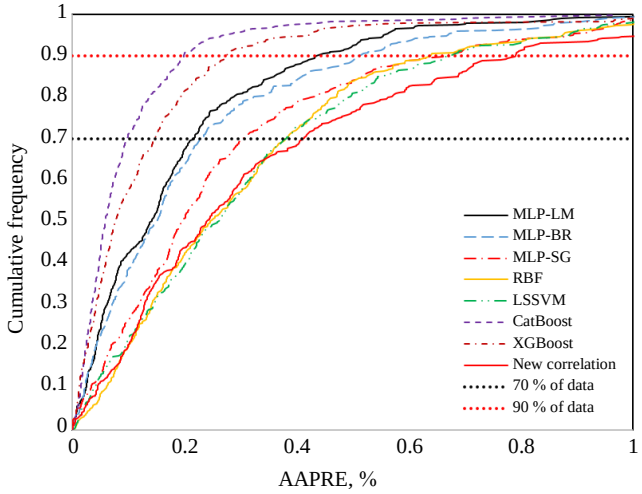


Fig.4. The cumulative frequency diagram of several ML models and the new correlation proposed in this survey

As noted above, the newly developed correlation model exhibited the highest performance among the proposed models, as illustrated in Fig.3.

The next step in visually evaluating the proposed ML models for predicting gas hydrate formation temperature is the analysis of the cumulative error frequency. This analysis allows the prediction of a larger proportion of data points with minimal errors. Figure 4 presents cumulative frequency distributions of AAPRE for various models applied to the entire dataset.

In Fig.4, the dark horizontal dashed line representing 70 % of the data indicates ARE, %: 0.1 – CatBoost, 0.15 – XGBoost, 0.2 – MLP-LM, 0.22 – MLP-BR, 0.29 – MLP-SCG, 0.37 – RBF, 0.38 – LSSVM, and 0.42 – new correlation models. These results demonstrate that 70 % of the dataset achieves prediction errors below the corresponding values. Additionally, approximately 90 % of the CatBoost model's predicted values exhibit absolute relative errors below 0.2 %, whereas the other models exceed this threshold. These findings, corroborated by statistical and graphical analyses, confirm the CatBoost model's superior accuracy in predicting gas hydrate formation temperature.

Trend analysis of the CatBoost and developed correlation. In this section, a trend analysis was conducted to examine whether the optimal models proposed in this study, namely CatBoost, and developed correlation accurately reflects the expected variations in gas hydrate formation temperature with respect to the selected independent variables. Figure 5 illustrate the relationship between gas hydrate formation temperature and pressure, while maintain a constant gas specific gravity.

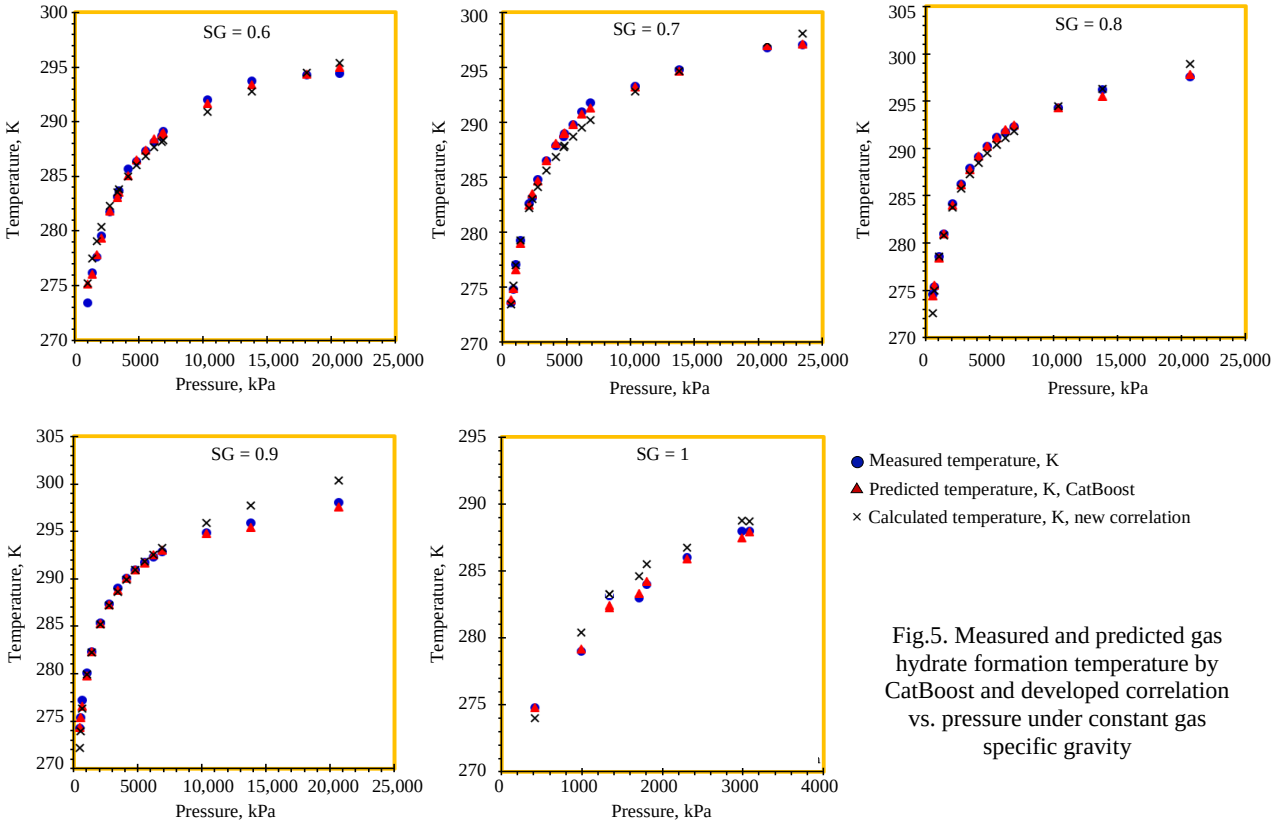


Fig.5. Measured and predicted gas hydrate formation temperature by CatBoost and developed correlation vs. pressure under constant gas specific gravity

This is an open access article under the CC BY 4.0 license

Figure 5 offers further confirmation of the accuracy of the developed models: for gas specific gravity values of 0.6, 0.7, 0.8, and 0.9, both the CatBoost model and the developed correlation closely match the experimental data. However, a slight discrepancy between the prediction models and the experimental data is observed at a gas specific gravity of 1. Furthermore, at high pressures across all gas specific gravities, the prediction accuracy of the models is slightly reduced. The primary reason for these minor deviations is attributed to the limited data available for a gas specific gravity of 1 and for higher pressures within the collected dataset. Additionally, experimental errors may also contribute to the discrepancies observed in the predictions of the developed correlation. Overall, as depicted in Fig.5, the predictions from both the CatBoost model and the developed correlation exhibit excellent agreement with the experimental data under various operational conditions. Additionally, Fig.5 demonstrates that, at a constant gas specific gravity, an increase in pressure leads to a higher gas hydrate formation temperature, a trend accurately captured by both the model and the correlation.

Sensitivity analysis. In this section of the present study, a sensitivity analysis using the absolute relevance factor *r* was performed to assess the influence of input parameters on gas hydrate formation temperature [67, 78, 79]. An input parameter is considered significant if its absolute relevance value is high. This factor is defined as follows:

$$r = \frac{abs\left(\sum_{i=1}^{n}\left(I_{k,i} - \bar{I}_k\right)\left(O_i - \bar{O}\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(I_{k,i} - \bar{I}_k\right)^2 \sum_{i=1}^{n}\left(O_i - \bar{O}\right)^2}},$$

where $I_{k,i}$, $\bar{I}_k$ – *k*-th input parameter and average value of I_k, respectively; *i* refers to either the specific pressure or pressure; O_i, $\bar{O}$ – output and the *i*-th average value of O_i.

The parameter *r*, ranging from –1 to 1, represents the correlation between the dependent and independent variables [80]. A positive value of *r* indicates a direct relationship, meaning that as the independent variable increases, the dependent variable also increases.

The sensitivity analysis of input variables affecting gas hydrate formation temperature, conducted using the Pearson correlation technique: gas specific gravity – 0.067; pressure – 0.752. Input variables such as pressure and gas specific gravity positively influence the target value, indicating that an increase in these parameters leads to an increase in gas hydrate formation temperature. Among them, pressure had the most significant impact, contributing 75.2 %, to the estimate of hydrate temperature. The gas hydrate temperature also increased with pressure.

Leverage approach. To evaluate the robustness of the proposed model and dataset, the Leverage method was applied. The proposed method aims to detect entries showing significant discrepancies across the data [81]. One of the key parameters of this method is the standardized residuals:

$$\mathrm{SR} = \frac{Z_i}{\left[\mathrm{MSE}\left(1 - H_{ii}\right)\right]^{0.5}},$$

where H_{ii} – hat indices; Z_i – error values; MSE – mean square error.

Another key parameter of this approach is the leverage threshold

$$H^* = \frac{3\left(I+1\right)}{N},$$

where *N* – total number of data points.

The last important parameter in this technique is the hat matrix

$$H = X\left(X^t X\right)^{-1} X^t,$$

where *X* – two-dimensional matrix; X^t – transpose [81-83].

This is an open access article under the CC BY 4.0 license

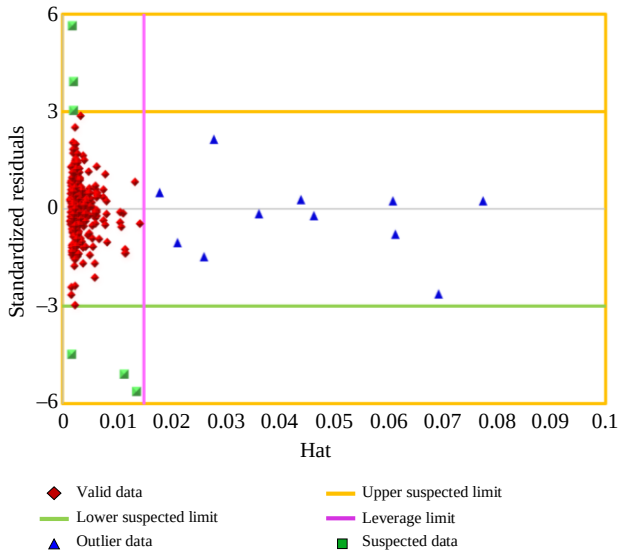


Fig.6. Williams plots for the complete dataset associated with the developed model in this study

William's plot demonstrates the correlation between standardized residuals SR and Hat values, dividing the data into four categories: bad leverage points, valid points, good leverage points, and upper or lower suspected points. A reliable model is characterized by having the majority of its data points fall in a range considered valid, with SR values ranging from –3 to 3 and Hat values staying below the leverage threshold. Data points are considered good leverage points when their Hat values exceed the threshold but their SR values remain within the range of –3 to 3, as they contribute to accurate predictions [71]. Conversely, leverage points that fall outside both domains are considered poor predictors. Suspected values are those with SR values outside the –3 to 3 range but with hat values below the leverage limit. A plot of SR against hat values was created for the CatBoost model to evaluate the formation temperature of gas hydrates (Fig.6).

Williams plot is used to assess the applicability scope of the new correlation and detect potential outliers. As shown in Fig.6, most data points fall within the ranges of $0 \leq H \leq 0.015$ and $-3 \leq \mathrm{SR} \leq 3$, indicating that the experimental data are of high quality and that the new developed model exhibits strong reliability. Data points with $\mathrm{SR} > 3$ or $\mathrm{SR} < -3$ are considered suspicious, because they suggest a high level of uncertainty. Outliers were identified as data points with $H > 0.015$ and $-3 < \mathrm{SR} < 3$, indicating that they fell outside the valid range of the proposed model. A total of 6 data points (1.3 % of the overall dataset) were identified as suspected outliers, with 11 outliers present in the temperature dataset, thereby reinforcing the reliability of the experimental dataset used for modeling.

Model limitations, applicable range, and future work. These models were developed using experimental data covering a pressure range of 330-68,600 kPa, gas specific gravity from 0.552 to 1.03, and laboratory temperatures between 272.04 and 304 K. Therefore, the proposed models are recommended for use within these specific ranges. Outside these boundaries, particularly at pressures lower than 330 kPa or higher than 68,600 kPa, temperatures outside the 272-305 K range, or for gas specific gravities not represented in the training data, prediction accuracy may decrease.

A slight discrepancy between the prediction models and the experimental data is observed at a gas specific gravity of 1 (Fig.5). Furthermore, at high pressures across all gas specific gravities, the prediction accuracy of the models is slightly reduced. The primary reason for these minor deviations is attributed to the limited data available for a gas specific gravity of 1 and for higher pressures within the collected dataset. The correlation is empirical and does not explicitly account for salinity, impurities, or kinetic non-equilibrium effects. These factors may introduce additional uncertainties under such conditions.

Conclusion

In this study, two distinct methodologies were employed to predict gas hydrate formation temperatures. The first approach was data-driven, utilizing several advanced machine learning algorithms: Multilayer Perceptron (MLP) optimized by Levenberg – Marquardt (LM), Bayesian Regularization (BR), and Scaled Conjugate Gradient (SCG); Radial Basis Function (RBF); and Least Squares Support Vector Machine (LSSVM). The second method involved developing a correlation using the

This is an open access article under the CC BY 4.0 license

Generalized Reduced Gradient (GRG) optimization technique, which accounts for variations in pressure and gas specific gravity. These models were developed and validated using a comprehensive dataset of 459 experimental temperature data points compiled from existing studies. Model performance was rigorously evaluated using statistical metrics, including the average absolute percentage relative error (AAPRE), root mean square error (RMSE), coefficient of determination (R^2), and further assessed through visual analyses such as cross-plots, error distribution graphs, and cumulative frequency distributions.

Among all models, the CatBoost algorithm demonstrated the highest predictive accuracy, achieving an AAPRE of 0.094 % and an R^2 value of 0.996. The subsequent models, XGBoost, MLP-LM, MLP-BR, MLP-SCG, RBF, and LSSVM, achieved AAPRE values of 0.103, 0.187, 0.209, 0.272, 0.308, and 0.311 %, respectively, indicating strong performance. The GRG-derived correlation also outperformed existing models in terms of predictive accuracy.

While the developed models exhibit high predictive accuracy within the training data's parameter space, their applicability is inherently limited by the range of input parameters used during development. Predictions made outside these established boundaries may result in reduced accuracy, a crucial consideration for practical implementation. Furthermore, the confidence in predictions can be affected by inherent uncertainties in the input data, particularly for highly relevant parameters. To provide a more critical accuracy assessment beyond standard metrics, we have analyzed outlier predictions, noting that they predominantly occur at the extremes of the input ranges or with less common data points. This offers a more nuanced understanding of the model's performance envelope. The comparative analysis against existing models has also been strengthened, clearly outlining scenarios where our models excel and acknowledging specific contexts where alternative approaches might be more suitable, thereby offering a balanced perspective relative to the state of the art. Sensitivity analysis revealed that pressure has a direct and positive influence on gas hydrate formation temperatures, with pressure being the most significant input variable.

Future research could expand the scope by incorporating influential factors such as salinity, gas composition, and the presence of inhibitors, which were not included in this study due to data limitations. Such an expansion would further enhance the generalizability and robustness of the predictive models.

Access to data

The full dataset used to predict the formation temperature of gas hydrate is available at the link: Supplement, https://pmi.spmi.ru/pmi/article/download/16663/16705.

REFERENCES

1. Safamirzaei M.P. Predict gas hydrate formation temperature with a simple correlation. *Gas Processing Journal*. 2015. July/August, p. 51-54.

2. Esfahani S., Baselizadeh S., Hemmati-Sarapardeh A. On determination of natural gas density: Least square support vector machine modeling approach. *Journal of Natural Gas Science and Engineering*. 2015. Vol. 22, p. 348-358. DOI: 10.1016/j.jngse.2014.12.003

3. Inkong K., Rangsunvigit P., Kulprathipanja S., Linga P. Effects of temperature and pressure on the methane hydrate formation with the presence of tetrahydrofuran (THF) as a promoter in an unstirred tank reactor. *Fuel*. 2019. Vol. 255. N 115705. DOI: 10.1016/j.fuel.2019.115705

4. Tatar A., Barati-Harooni A., Najafi-Marghmaleki A. et al. Prediction of moisture content of natural gases using a GA-RBF model. *Journal of Molecular Liquids*. 2016. Vol. 223, p. 994-999. DOI: 10.1016/j.molliq.2016.08.108

This is an open access article under the CC BY 4.0 license

5. Jingzhou Wang, Lu Zhang, Ran Li et al. A simple generalized viscosity model for natural gas components and their mixtures based on free-volume theory. *Journal of Molecular Liquids*. 2020. Vol. 320. Part B. N 114483. DOI: 10.1016/j.molliq.2020.114483

6. Jianbo Zhang, Zhiyuan Wang, Shun Liu et al. Prediction of hydrate deposition in pipelines to improve gas transportation efficiency and safety. *Applied Energy*. 2019. Vol. 253. N 113521. DOI: 10.1016/j.apenergy.2019.113521

7. Yan Wang, Koh C.A., White J. et al. Hydrate formation management simulations with anti-agglomerants and thermodynamic inhibitors in a subsea tieback. *Fuel*. 2019. Vol. 252, p. 458-468. DOI: 10.1016/j.fuel.2019.04.146

8. Zhiyuan Wang, Shikun Tong, Chao Wang et al. Hydrate deposition prediction model for deep-water gas wells under shut-in conditions. *Fuel*. 2020. Vol. 275. N 117944. DOI: 10.1016/j.fuel.2020.117944

9. Olajire A.A. Flow assurance issues in deep-water gas well testing and mitigation strategies with respect to gas hydrates deposition in flowlines – A review. *Journal of Molecular Liquids*. 2020. Vol. 318. N 114203. DOI: 10.1016/j.molliq.2020.114203

10. Chuanliang Yan, Yang Li, Yuanfang Cheng et al. Sand production evaluation during gas production from natural gas hydrates. *Journal of Natural Gas Science and Engineering*. 2018. Vol. 57, p. 77-88. DOI: 10.1016/j.jngse.2018.07.006

11. Kelland M.A., Iversen J.E. Kinetic Hydrate Inhibition at Pressures up to 760 Bar in Deep Water Drilling Fluids. *Energy & Fuels*. 2010. Vol. 24. Iss. 5, p. 3003-3013. DOI: 10.1021/ef9016152

12. Xin Zhao, Zhengsong Qiu, Zhen Zhang, Yubin Zhang. Relationship between the gas hydrate suppression temperature and water activity in the presence of thermodynamic hydrate inhibitor. *Fuel*. 2020. Vol. 264. N 116776. DOI: 10.1016/j.fuel.2019.116776

13. Yarveicy H., Ghiasi M.M. Modeling of gas hydrate phase equilibria: Extremely randomized trees and LSSVM approaches. *Journal of Molecular Liquids*. 2017. Vol. 243, p. 533-541. DOI: 10.1016/j.molliq.2017.08.053

14. Minwei Sun, Firoozabadi A. Gas hydrate powder formation – Ultimate solution in natural gas flow assurance. *Fuel*. 2015. Vol. 146, p. 1-5. DOI: 10.1016/j.fuel.2014.12.078

15. Sloan Jr. E.D., Koh C.A. Clathrate Hydrates of Natural Gases. CRC press, 2007, p. 752. DOI: 10.1201/9781420008494

16. Koh C.A., Westacott R.E., Zhang W. et al. Mechanisms of gas hydrate formation and inhibition. *Fluid Phase Equilibria*. 2002. Vol. 194-197, p. 143-151. DOI: 10.1016/S0378-3812(01)00660-4

17. Huangjing Zhao, Minwei Sun, Firoozabadi A. Anti-agglomeration of natural gas hydrates in liquid condensate and crude oil at constant pressure conditions. *Fuel*. 2016. Vol. 180, p. 187-193. DOI: 10.1016/j.fuel.2016.03.029

18. Li Wan, Na Zhang, De-Qing Liang. Inhibition effects of polysaccharides for gas hydrate formation in methane–water system. *Journal of Molecular Liquids*. 2019. Vol. 292. N 111435. DOI: 10.1016/j.molliq.2019.111435

19. Liang Yang, Zhenzhen Liu, Daoping Liu et al. Enhanced natural gas hydrates formation in the suspension with metal particles and fibers. *Journal of Molecular Liquids*. 2020. Vol. 301. № 112410. DOI: 10.1016/j.molliq.2019.112410

20. Hammerschmidt E. Formation of Gas Hydrates in Natural Gas Transmission Lines. *Industrial & Engineering Chemistry*. 1934. Vol. 26. Iss. 8, p. 851-855. DOI: 10.1021/ie50296a010

21. Kondori J., Zendehboudi S., James L. New insights into methane hydrate dissociation: Utilization of molecular dynamics strategy. *Fuel*. 2019. Vol. 249, p. 264-276. DOI: 10.1016/j.fuel.2019.02.125

22. Caili Dai, Yong Hu, Yining Wu et al. Effects of structural properties of alcohol molecules on decomposition of natural gas hydrates: A molecular dynamics study. *Fuel*. 2020. Vol. 268. N 117322. DOI: 10.1016/j.fuel.2020.117322

23. Bingbing Chen, Huiru Sun, Kehan Li et al. Experimental investigation of natural gas hydrate production characteristics via novel combination modes of depressurization with water flow erosion. *Fuel*. 2019. Vol. 252, p. 295-303. DOI: 10.1016/j.fuel.2019.04.120

24. Yutaek Seo, Seong-Pil Kang. Inhibition of methane hydrate re-formation in offshore pipelines with a kinetic hydrate inhibitor. *Journal of Petroleum Science and Engineering*. 2012. Vol. 88-89, p. 61-66. DOI: 10.1016/j.petrol.2011.11.001

25. Ju Dong Lee, Englezos P. Unusual kinetic inhibitor effects on gas hydrate formation. *Chemical Engineering Science*. 2006. Vol. 61. Iss. 5, p. 1368-1376. DOI: 10.1016/j.ces.2005.08.009

26. Bo-Hui Shi, Shang-Fei Song, Xiao-Fang Lv et al. Investigation on natural gas hydrate dissociation from a slurry to a water-in-oil emulsion in a high-pressure flow loop. *Fuel*. 2018. Vol. 233, p. 743-758. DOI: 10.1016/j.fuel.2018.06.054

27. Bahadori A. Fluid Phase Behavior for Conventional and Unconventional Oil and Gas Reservoirs. Gulf Professional Publishing, 2017, p. 554.

28. Soroush E., Mesbah M., Shokrollahi A. et al. Evolving a robust modeling tool for prediction of natural gas hydrate formation conditions. *Journal of Unconventional Oil and Gas Resources*. 2015. Vol. 12, p. 45-55. DOI: 10.1016/j.juogr.2015.09.002

29. Mesbah M., Soroush E., Rezakazemi M. Development of a least squares support vector machine model for prediction of natural gas hydrate formation temperature. *Chinese Journal of Chemical Engineering*. 2017. Vol. 25. Iss. 9, p. 1238-1248. DOI: 10.1016/j.cjche.2016.09.007

30. Berge B.K. Hydrate Predictions on a Microcomputer. Petroleum Industry Application of Microcomputers, 18-20 June 1986, Silvercreek, CO, USA. OnePetro, 1986. N SPE-15306-MS. DOI: 10.2118/15306-MS

31. Motiee M. Estimate possibility of hydrates. *Hydrocarbon Processing*. 1991. Vol. 70. N 7, p. 98-99.

32. Sun C.-Y., Chen G.-J., Lin W., Guo T.-M. Hydrate Formation Conditions of Sour Natural Gases. *Journal of Chemical & Engineering Data*. 2003. Vol. 48. Iss. 3, p. 600-602. DOI: 10.1021/je020155h

33. Towler B., Mokhatab S. Quickly estimate hydrate formation conditions in natural gases. *Hydrocarbon Processing*. 2005. Vol. 84. N 4, p. 61-63.

34. Salufu S.O., Nwakwo P. New Empirical Correlation for Predicting Hydrate Formation Conditions. SPE Nigeria Annual International Conference and Exhibition, 5-7 August 2013, Lagos, Nigeria. OnePetro, 2013. N SPE-167571-MS. DOI: 10.2118/167571-MS

35. Safamirzaei M. Predict gas hydrate formation temperature with a simple correlation. URL: https://gasprocessingnews.com/articles/2015/08/predict-gas-hydrate-formation-temperature-with-a-simple-correlation/ (accessed 09.02.2025).

36. Elgibaly A.A., Elkamel A.M. A new correlation for predicting hydrate formation conditions for various gas mixtures and inhibitors. *Fluid Phase Equilibria*. 1998. Vol. 152. Iss. 1, p. 23-42. DOI: 10.1016/S0378-3812(98)00368-9

37. Zahedi G., Karami Z., Yaghoobi H. Prediction of hydrate formation temperature by both statistical models and artificial neural network approaches. *Energy Conversion and Management*. 2009. Vol. 50. Iss. 8, p. 2052-2059. DOI: 10.1016/j.enconman.2009.04.005

38. Riazi S.H., Heydari H., Ahmadpour E. et al. Development of novel correlation for prediction of hydrate formation temperature based on intelligent optimization algorithms. *Journal of Natural Gas Science and Engineering*. 2014. Vol. 18, p. 377-384. DOI: 10.1016/j.jngse.2014.03.012

This is an open access article under the CC BY 4.0 license

39. Hesami S.M., Dehghani M., Kamali Z., Bakyani A.E. Developing a simple-to-use predictive model for prediction of hydrate formation temperature. *International Journal of Ambient Energy*. 2017. Vol. 38. Iss. 4, p. 380-388. DOI: 10.1080/01430750.2015.1100678

40. Amar M.N. Prediction of hydrate formation temperature using gene expression programming. *Journal of Natural Gas Science and Engineering*. 2021. Vol. 89. N 103879. DOI: 10.1016/j.jngse.2021.103879

41. Mesbah M., Habibnia S., Ahmadi S. et al. Developing a robust correlation for prediction of sweet and sour gas hydrate formation temperature. *Petroleum*. 2022. Vol. 8. Iss. 2, p. 204-209. DOI: 10.1016/j.petlm.2020.07.007

42. Hosseini M., Leonenko Y. A reliable model to predict the methane-hydrate equilibrium: An updated database and machine learning approach. *Renewable and Sustainable Energy Reviews*. 2023. Vol. 173. N 113103. DOI: 10.1016/j.rser.2022.113103

43. Karaköse M., Yücel Ö. Predictive Modeling of the Hydrate Formation Temperature in Highly Pressurized Natural Gas Pipelines. *Energies*. 2024. Vol. 17. Iss. 21. N 5306. DOI: 10.3390/en17215306

44. Khan N., Khan Ya., Zhang Lei et al. Investigating a User-Friendly Machine Learning Approach to Predict the Gas Hydrate Formation Temperature Available to Purchase. *Journal of the Geological Society of India*. 2026. Vol. 102 (4), p. 555-565. DOI: 10.17491/jgsi/2026/174383

45. Aleem W., Ahmad S., Qamar S. et al. Exploring Machine Learning Techniques for Accurate Prediction of Methane Hydrate Formation Temperature in Brine: A Comparative Study. *Arabian Journal for Science and Engineering*. 2024. Vol. 49. Iss. 6, p. 8983-8996. DOI: 10.1007/s13369-024-09030-5

46. Khafaie M., Masoumeh M., Azimi A., Mohammadi A. Comparing ANN with PSO-ANN in Prediction of Natural Gas Hydrate Formation Parameters. *Journal of the Japan Petroleum Institute*. 2025. Vol. 68. Iss. 2, p. 53-60. DOI: 10.1627/jpi.68.53

47. Mahmood Riyadh Atta, Akram Fadhl Al-Mahmodi, Bhajan Lal et al. Artificial Intelligence and Machine Learning in Thermodynamic Gas Hydrate Studies: A Review. *Energy & Fuels*. 2025. Vol. 39. Iss. 38, p. 18287-18310. DOI: 10.1021/acs.energyfuels.5c02863

48. Mokwenye P.O. Evaluation Of Gas Hydrate In Gas Pipeline Transportation: A Thesis Submitted to the Graduate Faculty of the University of North Dakota In partial fulfillment of the requirements for the degree of Master of Science. Grand Forks, 2020, p. 107. (N 3383). URL: https://commons.und.edu/theses/3383/ (дата обращения 09.02.2025).

49. Azizi J., Azizi J., Davarnejad R. Prediction of Gas Hydrate Formation using HYSYS Software. *International Journal of Engineering, Transactions C: Aspects*. 2014. Vol. 27. Iss. 9, p. 1325-1330.

50. Bahadori A., Vuthaluru H.B. A novel correlation for estimation of hydrate forming condition of natural gases. *Journal of Natural Gas Chemistry*. 2009. Vol. 18. Iss. 4, p. 453-457. DOI: 10.1016/S1003-9953(08)60143-7

51. Wilcox W.I., Carson D.B., Katz D.L. Natural Gas Hydrates. *Industrial & Engineering Chemistry*. 1941. Vol. 33. Iss. 5, p. 662-665. DOI: 10.1021/ie50377a027

52. Guang-Jin Chen, Tian-Min Guo. Thermodynamic modeling of hydrate formation based on new concepts. *Fluid Phase Equilibria*. 1996. Vol. 122. Iss. 1-2, p. 43-65. DOI: 10.1016/0378-3812(96)03032-4

53. Jøssang A., Stange E. A new predictive activity model for aqueous salt solutions. *Fluid Phase Equilibria*. 2001. Vol. 181. Iss. 1-2, p. 33-46. DOI: 10.1016/S0378-3812(00)00515-X

54. Sloan E.D. Clathrate hydrates of natural gases. Marcel Dekker, New York, 1990, p. 641.

55. Nasrifar K., Moshfeghian M. A model for prediction of gas hydrate formation conditions in aqueous solutions containing electrolytes and/or alcohol. *The Journal of Chemical Thermodynamics*. 2001. Vol. 33. Iss. 9, p. 999-1014. DOI: 10.1006/jcht.2000.0811

56. Deaton W.M., Frost E.M. Gas hydrates and their relation to the operation of natural-gas pipe mines. American Gas Association. United States. Bureau of Mines. Monograph 8, p. 101.

57. Hashemkhani M., Soleimani R., Fazeli H. et al. Prediction of the binary surface tension of mixtures containing ionic liquids using Support Vector Machine algorithms. *Journal of Molecular Liquids*. 2015. Vol. 211, p. 534-552. DOI: 10.1016/j.molliq.2015.07.038

58. Suykens J.A.K., Van Gestel T., De Brabanter J. et al. Least Squares Support Vector Machines. World Scientific, 2002, p. 308. DOI: 10.1142/5089

59. Suykens J.A.K., Vandewalle J. Least Squares Support Vector Machine Classifiers. *Neural Processing Letters*. 1999. Vol. 9. Iss. 3, p. 293-300. DOI: 10.1023/A:1018628609742

60. Suykens J.A.K., De Brabanter J., Lukas L., Vandewalle J. Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing*. 2002. Vol. 48. Iss. 1-4, p. 85-105. DOI: 10.1016/S0925-2312(01)00644-0

61. Varamesh A., Hemmati-Sarapardeh A., Keshavarz Moraveji M., Mohammadi A.H. Generalized models for predicting the critical properties of pure chemical compounds. *Journal of Molecular Liquids*. 2017. Vol. 240, p. 777-793. DOI: 10.1016/j.molliq.2017.05.023

62. Varamesh A., Hemmati-Sarapardeh A., Dabir B., Mohammadi A.H. Development of robust generalized models for estimating the normal boiling points of pure chemical compounds. *Journal of Molecular Liquids*. 2017. Vol. 242, p. 59-69. DOI: 10.1016/j.molliq.2017.06.039

63. Ningbo Zhao, Xueyou Wen, Jialong Yang et al. Modeling and prediction of viscosity of water-based nanofluids by radial basis function neural networks. *Powder Technology*. 2015. Vol. 281, p. 173-183. DOI: 10.1016/j.powtec.2015.04.058

64. Faris H., Aljarah I., Mirjalili S. Evolving Radial Basis Function Networks Using Moth–Flame Optimizer. Handbook of Neural Computation. Elsevier, 2017, p. 537-550. DOI: 10.1016/B978-0-12-811318-9.00028-4

65. Hemmati-Sarapardeh A., Varamesh A., Husein M.M., Karan K. On the evaluation of the viscosity of nanofluid systems: Modeling and data assessment. *Renewable and Sustainable Energy Reviews*. 2018. Vol. 81. Part 1, p. 313-329. DOI: 10.1016/j.rser.2017.07.049

66. Turnbull D., Elkan C. Fast recognition of musical genres using RBF networks. *IEEE Transactions on Knowledge and Data Engineering*. 2005. Vol. 17. Iss. 4, p. 580-584. DOI: 10.1109/TKDE.2005.62

67. Mehrjoo H., Riazi M., Nait Amar M., Hemmati-Sarapardeh A. Modeling interfacial tension of methane-brine systems at high pressure and high salinity conditions. *Journal of the Taiwan Institute of Chemical Engineers*. 2020. Vol. 114, p. 125-141. DOI: 10.1016/j.jtice.2020.09.014

68. Riazi M., Mehrjoo H., Nait Amar M. et al. Modeling the density of acid gases at extensive ranges of pressure and temperature conditions. *Journal of Petroleum Science and Engineering*, 2021. Vol. 207. N 109063. DOI: 10.1016/j.petrol.2021.109063

69. Rostami A., Hemmati-Sarapardeh A., Shamshirband S. Rigorous prognostication of natural gas viscosity: Smart modeling and comparative study. *Fuel*. 2018. Vol. 222, p. 766-778. DOI: 10.1016/j.fuel.2018.02.069

70. Zendehboudi A., Tatar A., Xianting Li. A comparative study and prediction of the liquid desiccant dehumidifiers using intelligent models. *Renewable Energy*. 2017. Vol. 114. Part B, p. 1023-1035. DOI: 10.1016/j.renene.2017.07.078

This is an open access article under the CC BY 4.0 license

71. Riazi M., Mehrjoo H., Nakhaei R. et al. Modelling rate of penetration in drilling operations using RBF, MLP, LSSVM, and DT models. *Scientific Reports*. 2022. Vol. 12. N 11650. DOI: 10.1038/s41598-022-14710-z

72. Dabiri M.-S., Hadavimoghaddam F., Ashoorian S. et al. Modeling liquid rate through wellhead chokes using machine learning techniques. *Scientific Reports*. 2024. Vol. 14. N 6945. DOI: 10.1038/s41598-024-54010-2

73. Xinyuan Zou, Yingting Zhu, Jing Lv et al. Toward Estimating CO_2 Solubility in Pure Water and Brine Using Cascade Forward Neural Network and Generalized Regression Neural Network: Application to CO_2 Dissolution Trapping in Saline Aquifers. *ACS Omega*. 2024. Vol. 9. Iss. 4, p. 4705-4720. DOI: 10.1021/acsomega.3c07962

74. Nanda A.K., Neelakandan S., Gupta S. et al. Multi-layer perceptron's neural network with optimization algorithm for greenhouse gas forecasting systems. *Environmental Challenges*. 2023. Vol. 11. N 100708. DOI: 10.1016/j.envc.2023.100708

75. MacKay D.J.C. Bayesian Interpolation. *Neural Computation*. 1992. Vol. 4. Iss. 3, p. 415-447. DOI: 10.1162/neco.1992.4.3.415

76. Makomere R., Rutto H., Koech L., Banza M. The use of artificial neural network (ANN) in dry flue gas desulphurization modelling: Levenberg–Marquardt (LM) and Bayesian regularization (BR) algorithm comparison. *The Canadian Journal of Chemical Engineering*. 2023. Vol. 101. Iss. 6, p. 3273-3286. DOI: 10.1002/cjce.24715

77. Chavoshi S., Safamirzaei M., Pajoum Shariati F. Evaluation of Empirical Correlations for Predicting Gas Hydrate Formation Temperature. *Gas Processing Journal*. 2018. Vol. 6. Iss. 2, p. 15-36. DOI: 10.22108/gpj.2018.112052.1036

78. Shateri M., Ghorbani S., Hemmati-Sarapardeh A., Mohammadi A.H. Application of Wilcoxon generalized radial basis function network for prediction of natural gas compressibility factor. *Journal of the Taiwan Institute of Chemical Engineers*. 2015. Vol. 50, p. 131-141. DOI: 10.1016/j.jtice.2014.12.011

79. Benamara C., Gharbi K., Nait Amar M., Hamada B. Prediction of Wax Appearance Temperature Using Artificial Intelligent Techniques. *Arabian Journal for Science and Engineering*. 2020. Vol. 45. Iss. 2, p. 1319-1330. DOI: 10.1007/s13369-019-04290-y

80. Castillo E., Hadi A.S., Conejo A., Fernández-Canteli A. A General Method for Local Sensitivity Analysis With Application to Regression Models and Other Optimization Problems. *Technometrics.* 2004. Vol. 46. Iss. 4, p. 430-444. DOI: 10.1198/004017004000000509

81. Rousseeuw P.J., Leroy A.M. Robust Regression and Outlier Detection. John Wiley & Sons, 2003, p. 360.

82. Mehrjoo H., Riazi M., Rezaei F. et al. Modeling interfacial tension of methane-brine systems at high pressure high temperature conditions. *Geoenergy Science and Engineering*. 2024. Vol. 242. N 213258. DOI: 10.1016/j.geoen.2024.213258

83. Dabiri M.-S., Haji-Hashemi R., Ashoorian S. et al. Modeling drilling fluid density at high-pressure high-temperature conditions using advanced machine-learning techniques. *Geoenergy Science and Engineering*. 2025. Vol. 244. N 213369. DOI: 10.1016/j.geoen.2024.213369

***Authors*: Mohsen Riazi,** *MSc* (*Shahid Bahonar University of Kerman, Kerman, Iran*), **Hossein Mehrjoo,** *PhD* (*Shahid Bahonar University of Kerman, Kerman, Iran*), **Mohammad-Saber Dabiri,** *MSc* (*Shahid Bahonar University of Kerman, Kerman, Iran*), **Jamilyam Ismailova,** *PhD, Associate Professor* (*Satbayev University, Almaty, Republic of Kazakhstan*), **Masoud Riazi,** *PhD, Associate Professor* (*Nazarbayev University, Astana, Republic of Kazakhstan*), *masoud.riazi@nu.edu.kz*.

The authors declare no conflict of interests.

This is an open access article under the CC BY 4.0 license