Journal of Environmental Economics & Chemical Processes (JEECP)

https://doi.org/10.22034/jeecp.2026.2088082.1018

Journal Homepage: *WWW.JEECPJournal.com*

Journal of Environmental Economics & Chemical Processes (JEECP)

RIGINAL RESEARCH PAPER

Modeling and Simulation of CO_2 and CH_4 Adsorption Behavior on Shale Rocks Using Data Science Approach: An Insight into CO_2 Sequestration and Shale Gas Production

Abbas Ayatizadeh [a]; Ghasem Zargar [b]; Touba Hamoule [c, *]

[a] *Ahwaz Faculty of Petroleum, Petroleum University of Technology, Abadan, Iran.*
[b] *Department of Petroleum Engineering, Ahwaz Faculty of Petroleum, Petroleum University of Technology, Abadan, Iran*
[c] *Department of Basic Science, Ahwaz Faculty of Petroleum, Petroleum University of Technology, Abadan, Iran.*

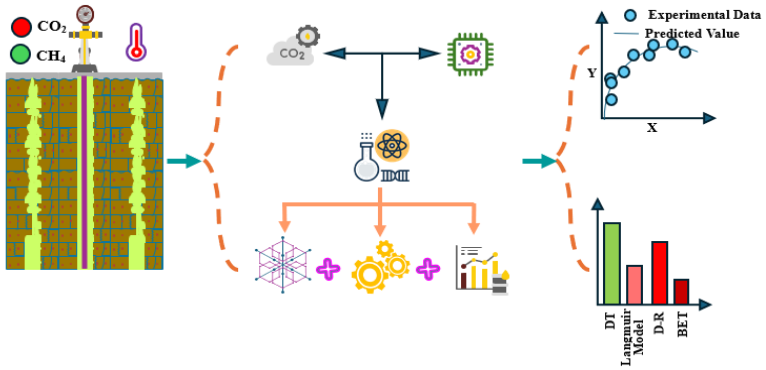
P A P E R I N F O

Paper history:
Received: 06/01/2026
Revised: 06/25/2026
Accepted in revised form: 06/29/2026
Published: 06/30/2026

Keywords:
CO_2 Storage;
Methane Recovery;
Shale Rocks;
Data Science;
CO_2 Adsorption.

A B S T R A C T

Shale gas reservoirs are suitable for CO_2 sequestration to reduce CO_2 emissions. In Given the uncertainties and complexities associated with determining the adsorption capacity of CO_2 in various shale rocks, developing a generalized and comprehensive model for precise CO_2 adsorption estimates on shale rocks at different conditions is crucial and of high importance. In this study, data science approach is used for developing accurate models for the estimation of CO_2 and CH_4 adsorption on shale samples with different properties and at various ratios of CO_2 and CH_4 in the mixture. A comprehensive data bank including 2542 data points of both excessive and absolute adsorption of CO_2 and CH_4 adsorption on shale rocks entailing pressure, temperature, CO_2 and CH_4 ratios, and total organic carbon (TOC) has been gathered and used for model development. seven intelligent models namely Least Square Support Vector Machine, K-nearest Neighbors, Artificial Neural Network, Decision Tree, Adaptive Boosting, Random Forest, and Ensemble Learning were utilized to estimate adsorption capacity. The accuracy and performance of the integrated models were evaluated and compared through statistical and graphical analyses. The obtained results were also compared with well-established adsorption models including Langmuir, Brunauer-Emmett-Teller (BET), Dubinin-Astakhov (D-A), and Dubinin-Radushkevich (D-R). The findings showed that the decision tree model has the highest accuracy with the coefficient of determination (R^2) of 0.999 and average absolute relative error (AARE %) of 7.138 for total data. In addition, all the developed intelligent models exhibit higher accuracy than the models found in the literature. The D-R model with R^2 and AARE % of 0.987 and 31.319, respectively, is the worst model. Moreover, sensitivity analysis was performed in order to ascertain the effect of each influencing parameter on the adsorption capacity of shale rocks. It is shown that pressure and TOC are the most effective factors on the output variable..



1. Introduction

Global warming and climate change are primarily caused by the greenhouse effect. The storage of CO_2 gas has been considered as a key factor in reducing greenhouse gases [1]. For the storage of CO_2, shale gas reserves may be viable alternatives within productive shale gas reservoirs[2]. These reservoirs are considered as excellent options for carbon dioxide storage due to their fine porous space, effective sealing capabilities, and the ability to replace methane with carbon dioxide through strong adsorption in shale [3-5]. Additionally, this storage solution provides a means for oil and gas recovery while mitigating the high costs associated with storage.

* Corresponding Author: Touba Hamoule
E-mail address: t.hamoule@put.ac.ir
ISSN: 3116-0050
URL: https://www.jeecpjournal.com/article_735392.html

Please cite this article as: : A. Ayatizadeh, M. Zargar, T. Hamoule *Journal of Enviromental Economics & Chemical Processes (JEECP)*, 2026, *3*(2), 75-85. https://doi.org/10.22034/jeecp.2026.2088082.1018

/© 2026 The Author(s). Published by Power of New Vision Movement Institute.

The adsorption of carbon dioxide (CO_2) onto shale surfaces plays a critical role during the energy extraction process from unconventional reservoirs, particularly in the context of enhanced gas recovery (EGR) and geological CO_2 sequestration. This phenomenon is not only a fundamental aspect of gas–shale interactions but also has significant implications for both reservoir performance and environmental sustainability. The coexistence of CO_2 adsorption with energy extraction processes opens up opportunities for dual-purpose operations: CO_2 Storage and Enhanced Methane Recovery. However, the success of such integrated approaches depends on several factors including competitive adsorption behavior of CO_2 and CH_4, pressure and temperature conditions, organic richness and mineralogy of the shale, and pore structure and accessibility. The adsorption of CO_2 during the energy extraction process from shale reservoirs is a complex yet strategically important phenomenon. Understanding and leveraging this interaction can lead to more sustainable and efficient energy recovery strategies, aligning fossil fuel extraction with global efforts toward carbon neutrality [6-8].

Potentially increasing CH_4 recovery can be achieved by injecting CO_2 into shale reservoirs. Additionally, the recovered CH_4 might reduce the cost of the CO_2 sequestration procedure [9,10]. At a specific pressure and temperature, the organic matter content of shales with substantial organic matter affects their capacity to adsorb gases. Total organic content (TOC), according to exploratory research by Mani et al., (2017) [11], played a first-order impact on the adsorptive capacity. Numerous researchers have conducted experiments and numerical simulation studies in the field of carbon storage within shale gas formations. In real shale samples as well as the pure mineral components, Heller and Zoback, (2014) [12] demonstrated that carbon dioxide has an adsorptive capacity that is roughly 2-3 times greater than that of methane. Additionally, they were able to correctly fit the Langmuir isotherm model to all of the adsorption data. Merey and Sinayuc, (2018) [13] conducted CH_4 and CO_2 adsorption experiments on shale samples, at temperatures of 25°C, 50°C, and 75°C. The primary goal was to assess CO_2 injection as an enhanced recovery technique for shale gas reservoirs. Langmuir and Ono-Kondo models were used to analyze the adsorption data, and layered structures of adsorption were explored. These models were employed to evaluate the adsorption data and understand the layered structures of adsorption on the shale samples. Du et al., (2021) [14] utilized an improved experimental procedure and data processing method proposed in their work. They demonstrate that the quantities of recovered CH_4 and stored CO_2 rise in conjunction with an increase in CO_2 injection pressure. Liao et al., (2023) [9] Used volumetric method to calculate the adsorption behavior of pure and binary CH_4/CO_2 in shale at a higher pressure of 13 MPa. Shale preferentially absorbed CO_2, and the amount of binary CO_2/CH_4 that was adsorbed increased as the CO_2 content increased, according to pure and binary adsorption experiments. Qin et al., (2020) [15] investigated the impact that supercritical CO_2 ($ScCO_2$) exposure has on the kinetics of CH_4 and CO_2 adsorption on shale. On both untreated and $ScCO_2$-treated shale samples, the researchers performed X-ray diffraction analyses, low-pressure N_2 adsorption, and adsorption kinetics experiments. The adsorption rates of CH_4 and CO_2 were fitted using several adsorption kinetics models, and the Bangham adsorption (BA) model demonstrated the best-fitting effect ($R^2 > 0.99$). An appropriate model is required to predict the adsorption on a variety of shales, but it is difficult to construct a universal relationship between the characteristics and adsorption behavior of CO_2/CH_4 due to the high complexity of CH_4 and CO_2 adsorption. The CO_2/CH_4 binary mixture's competitive adsorption features and process were examined by Zhou et al., (2019) [16], and the key element dictating the behavior of competitive adsorption was identified. Few studies concentrated on the parameters of each component, while the majority concentrated on the overall reservoir pressure and temperature characteristics. Accurate experimental data are needed to create mathematical models for adsorption that are credible. Such results have been obtained experimentally using chromatographic, gravimetric, and volumetric methods. Li et al., (2022) [17] investigates the particular adsorption behaviors of CO_2/CH_4 in clay-mineral nanopores utilizing the grand canonical Monte Carlo (GCMC) simulation method. Following the experimental adsorption data, several adsorption isotherm models have been proposed. The Langmuir, D-P (Dubinin-Polanyi), and BET (Brunauer-Emmett-Teller) models can be used to extract absolute adsorption parameters from methane or CO_2 experimental adsorption isotherm data [18]. Although these equations were established for various adsorbent-adsorbate systems, they were also shown to be applicable for coal and shale gas systems additionally.

In order to model CO_2 adsorption, Clarkson and Haghshenas, (2013) [19] examined the applicability of numerous models. In order to predict supercritical single-component adsorption on coal and shale, they found that the straightforward Langmuir and Dubinin-Radushkevich models were the most effective. Li et al., (2022) [17] with correlation coefficients ranging from 0.9896 to 0.9976, the Onoe-Kondo model was employed to fit the measured methane adsorption isotherms. The model described the filling adsorption characteristics of nanopores in supercritical conditions.

Nevertheless, experimental adsorption studies are time-consuming and costly, as well as are not always accessible for shale reservoirs of interest or under the studied geologic conditions. As a result, new methodologies for estimating carbon dioxide adsorption on the shale surface are required.[20, 21]. Multiple investigations have revealed that the Langmuir model fits methane adsorption data in shales effectively [22]. Conventional models, like the Langmuir model, are not generally applicable since they depend extensively on complicated experimenting. To determine CO_2/CH_4 competitive adsorption ratios—which are essential for CO_2 sequestration and natural gas extraction from shale resources—a unified adsorption model must be developed. Only a few empirical models with inadequate precision exist for measuring the adsorbed gas, whereas there are several precise models for calculating the free gas inside the pore space of shales. Adsorption research is also difficult, time-consuming, and subject to several error causes, such as gas loss during sample preparation and coring. Advances in data science and machine learning techniques have significantly improved computing efficiency and prediction accuracy in recent years.[23-27]. These machine-learning approaches are currently employed in various fields of research. Only a few research have utilized intelligent modeling approaches to forecast methane adsorption capacity.

Ma et al., (2023) [28] performed a comparative machine learning method analysis to accurately predict the highest possible CO_2 adsorption capacity and CO_2/CH_4 effective adsorption ratios. The ANN, XGBoost, and RF approaches were examined together with four critical input parameters (temperature, TOC, humidity, and maximum adsorption capacity of CH_4). It was found that the ANN models can effectively expect the required outputs within a 10% accuracy level by assessing the mean absolute errors (MAE) and R^2. Amar et al., (2022) [29] employed two rigorous data-driven techniques, group method of data handling (GMDH) and gene expression programming (GEP), to provide reliable and consistent explicit mathematical expressions for predicting methane adsorption. A full database with 352 data points was built for this reason from the literature. The developed correlations used pressure, temperature, moisture, and total organic carbon (TOC) as input variables. Results indicate that both correlations are capable of making accurate estimations. With a correlation value of 0.9837, the GEP-based correlation, however, provides more accurate predictions for methane adsorption. In additional work, Bemani et al., (2020) [30] employed the LSSVM algorithm to estimate the CO_2 and CH_4 adsorption capacity in shale while taking into account TOC, temperature, pressure, and gas mole fraction of CO_2 as input variables. They optimized the LSSVM model using the particle swarm optimization (PSO) technique, and they reported that the R^2 of the testing dataset for methane adsorption was 0.9966.

With the goal of identifying an effective data-driven model for predicting the competitive adsorption ratios of CO_2/CH_4 for shale gas reservoirs, a systematic comparison research was carried out in this research based on the experimental data gathered from previous research. The aim of this article is to model the adsorption of CO_2 and its compounds with CH_4 on Shale. For this modeling, data-based models or artificial intelligence models were employed, and the results of these models were compared with conventional adsorption models. Artificial intelligence methods employed in the study include ANN, SVM, DT, KNN, EL, FR, and Adaptive Boosting. Additionally, conventional adsorption models such as Langmuir, BET model, D-A model, and D-R model were utilized. About 2542 data points from various sources, related to CO_2 adsorption and CO_2-CH_4 mixtures on all types of shales, were collected for this modeling. The model inputs consist of temperature, pressure, the ratio of CO_2 to CH_4, and the TOC of the shale. From this model, two types of adsorption parameters can be obtained: absolute adsorption and excessive adsorption.

2. Method

2.1 Artificial neural network

ANNs is a kind of machine learning algorithm that is based on the composition and operations of the human brain. ANNs consist of interconnected processing nodes, called artificial neurons, organized into layers that receive inputs, perform calculations, and pass outputs to the next layer. [31]. They are trained using backpropagation, which involves adjusting the weights of connections between neurons based on the error between the network's output and the desired output, as shown in Figure 1 [32]. It has been demonstrated that ANNs are highly effective at solving complex problems that are difficult to solve using traditional programming techniques [33]. An ANN is a multilayer neural network composed of neuron units that are completely interconnected with neuron units in neighboring layers (i.e., input layer, hidden layer, and output layer) without interconnecting with neurons in the same layer. An ANN functions by accepting limited external input, learning intricate correlations between data samples using specific algorithms, and dynamically adjusting to patterns in the data[34]. The cost or error function, which offers a measure of variation between the expected result and the intended output, provides the basis of the learning algorithm. To identify the most effective solution, the MSE function, which represents this cost function, must typically be reduced. MSE has two benefits: it offers an instantaneous learning process and is the most suitable cost function for the learning algorithm [35]. MSE is provided by:

$$MSE = \frac{1}{n_{OP} \cdot n_{dp}} \sum_{j=1}^{j=n_{OP}} \sum_{i=1}^{i=n_{dp}} \left(t_{ij} - y_{ij} \right)^2 \quad (1)$$

2.2 Least-squares support vector machine

LSSVM is a modified version of the SVM that replaces inequality constraints with equality constraints. It models hysteresis by estimating the autoregressive function and utilizes a nonlinear function to transform the input into a higher-

dimensional space [36]. LSSVM is applied in hysteresis modeling and offers a fitting capability comparable to that of artificial neural networks (ANN) [37]. The key advantage of LSSVM is its computational efficiency, as it transforms the quadratic programming problem of standard SVMs into a set of linear equations. This makes it particularly effective for large-scale problems [30].

2.3. Decision tree

The supervised learning class of algorithms, which includes the Decision Tree (DT) method, is mostly used to solve classification issues, while it may also be used to solve regressive problems [38]. A node is made up of a leaf node that shows the result, an inner node that shows the branch structures, and a dataset that describes the algorithm's decision. There are two nodes: the decision node, which makes choices and has many branches, and the leaf node, which derives from decision nodes and has no more branches [39]. Due to the similarity in form, it shares a name with a tree. The structure resembles a tree since the root node acts as a starting point and subsequently expands to several branches, as shown in Figure 3. Depending on whether the answer to the query is yes or no, the decision tree splits into sub-trees. [40].

2.4. K-Nearest Neighbor

KNN was initially created by Cover and Hart for classification tasks. It has been extensively utilized to carry out non-parametric regressions throughout the last few decades. State vectors are built using past data (the dataset) and current data (new unobserved data) to determine the goal value of the KNN algorithm [41]. The k-nearest neighbors in this approach are the k data points having the shortest distance (Euclidean Distance in this study) between them. By computing the average value of k neighbors for each new data point, it is feasible to obtain the prediction results for the target value. Figure 4 provides a thorough description of KNN. The complexity of this method in terms of time is O(n) [42].

2.5. Ensemble Learning

EL is the deliberate creation and fusion of many models, such as classifiers or experts, to address a particular computational intelligence challenge[43, 44]. EL is usually applied to improve a model's performance (prediction, classification, function approximation, etc.) or lessen the possibility that an unsuitable model will be inadvertently selected. [45]. Additional applications of ensemble learning include giving the model's conclusion a level of confidence, nonstationary learning, data fusion, choosing the best features, error correction, and incremental learning [46]. Boosted trees, as seen in Figure 5, were the model that the EL algorithm found to be the most accurate fit.Figure 4 provides a thorough description of K-Nearest Neighbor.

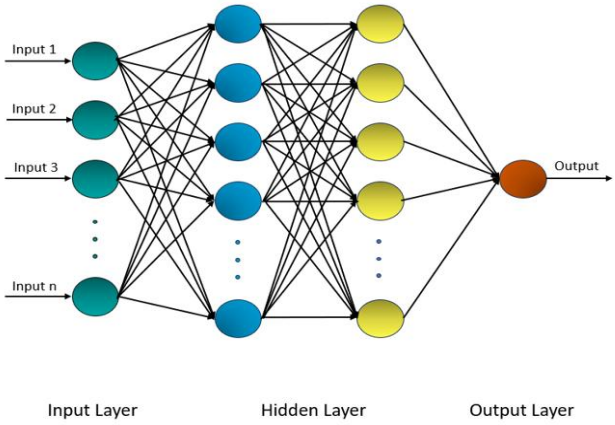


Figure 1. Illustration of an Artificial Neural Network

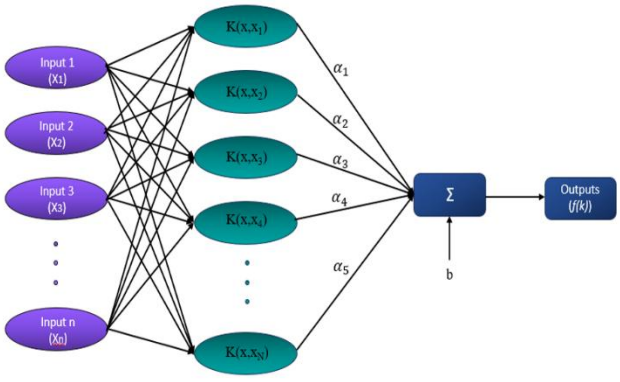


Figure 2. The associated structure of the LSSVM model

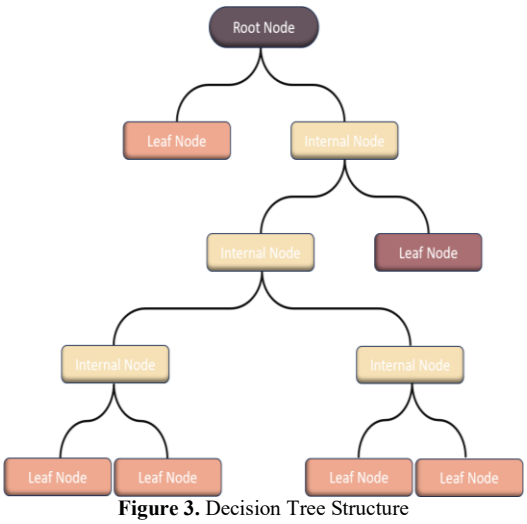


Figure 3. Decision Tree Structure

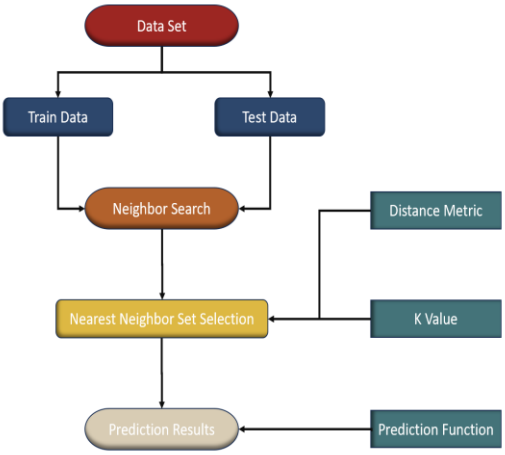


Figure 4. The procedure of KNN Regression model

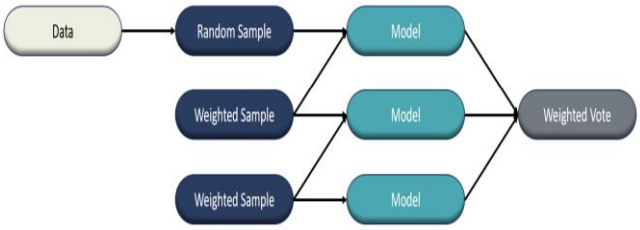


Figure 5. Illustration of Ensemble Learning

2.6. Random Forest

The RF is an ensemble learning algorithm that makes use of Berum's original bagging method. To improve prediction accuracy, RF integrates forecasts from many decision trees. The result variation from individual decision trees is reduced by RF's use of the bootstrap and aggregation approach [38]. The RF method randomly chooses the characteristics for each tree. Then, regression and classification tasks are carried out using the gathered outputs from multiple trees. For classification tasks, the outcomes of the grouped learners are voted on, but for regression problems, the results are averaged [47]. RF frequently resolves under-fitting concerns, which frequently arise in a single decision tree approach, extremely successfully in a low-dimensional dataset [48].

2.7. Adaptive Boosting

The literature has a variety of boosting algorithm kinds. The AdaBoost or Adaptive Boosting algorithm is the most popular type of boosting algorithm that is often used in practice. Freund and Schapire created the AdaBoost algorithm in 1995 [49]. With this approach, prior knowledge of how the weak regressor/classifier procedure works is not necessary. Additionally, this approach may be used with several weak learner types, including CART and SVM. Another benefit is that when used in conjunction with a suitable weak learner, AdaBoost is not susceptible to the overfitting problem. In fact, overfitting may happen if the employed weak learner is too complicated or too weak. The AdaBoost algorithm has this drawback [50].

2.8. Langmuir

The Langmuir adsorption isotherm is an equilibrium model that depicts the monolayer maximum coverage of gas on the rock surface, assuming homogenous and equal adsorption sites. Eq. 4 represents the Langmuir isotherm, which is derived from a localized adsorption model [51]. This is the most fundamental and widely used isotherm for gas adsorption on homogeneous surfaces. It has been extensively applied to model CH_4 and CO_2 adsorption in shale and coal systems. Including it provides a baseline for evaluating whether more complex models are necessary [52].

$$\frac{q}{Q_l} = \frac{k_l P}{1 + k_l P} \tag{4}$$

The model parameters (k_l and Q_l) are determined from the slope and intercept of a plot of (1/q) versus (1/p).[53].

$$\frac{1}{q} = \frac{1}{Q_l} + \frac{1}{Q_l k_l}\frac{1}{P} \tag{5}$$

However, because rock surfaces are naturally heterogeneous, the Langmuir isotherm typically fails short of capturing the characteristics of rock adsorption.

2.9. Brunauer-Emmett-Teller (BET)

By modifying the Langmuir model Eq.5, [54] created the BET model. The BET model describes both monolayer and multilayer adsorption processes, despite being often utilized to address questions with multilayer adsorption. Unlike Langmuir, the BET model accounts for multilayer adsorption, which is particularly relevant at higher pressures where additional molecular layers may form on the shale surface [55].

$$\frac{Q_{abs}}{Q_m} = \frac{c\left(\frac{P}{P_0}\right)}{\left(1-\frac{P}{P_0}\right)\left[1-(1-c)\left(\frac{P}{P_0}\right)\right]} \tag{6}$$

2.10. Dubinin and Radushkevich (DR)

Based on the theory that the pores in a heterogeneous adsorbent follow the micropore volume-filling process and have a Gaussian energy distribution, the DR model was developed. Shale rocks are known to contain significant microporosity, making the D-R model highly relevant for describing adsorption in micropores [56, 57]. The basic formula is as below:

$$Q_{abs} = Q_0 \exp\left\{-D\left[\ln\left(\frac{P_0}{P}\right)\right]^2\right\} \tag{7}$$

2.11. Dubinin–Astakhov (D-A)

The D-A model has been shown to provide superior fits for adsorption on heterogeneous shale surfaces, particularly when the adsorbent exhibits a wide range of pore sizes and surface chemistries.[58] The D-A model Eq. (7) modifies the D-R model and adds a fitting parameter, greatly increasing flexibility. where a represents the shale's structural heterogeneity constant, with values in the 1- 4 range [59].

$$Q_{abs} = Q_m \exp\left\{-D\left[\ln\left(\frac{P_0}{P}\right)\right]^a\right\} \tag{7}$$

Developing an accurate and efficient model needs comprehensive and reliable experimental data from different resources. In this study, an extensive dataset including 2542 data points of CO_2/CH_4 adsorption on different types of shale rocks was gathered from various references [16, 23, 60-64] for developing data-driven and empirical models. A precise modeling approach heavily depends on the quality of the data employed. Therefore, preprocessing analyses including removing duplicates and low-variance features, collinearity assessment, outlier removal, etc. were performed on the gathered data bank. Moreover, feature selection is one of the most important parts of a modeling study. Parameters including pressure, temperature, CO_2 and CH_4 concentration, shale characteristics, and type of adsorption are the main features that should be considered in the developed model. Shale characteristics contain the TOC content, mineral composition, average pore diameter (APD), specific surface area (SSA), etc. Due to the lack of data, the TOC content as a representative of the shale characteristics was considered in the developed models. TOC content has shown to be an effective factor in adsorption amount of both CO_2 and CH_4 (Ma et al., 2019; Zhou et al., 2019). The collected dataset includes adsorption data at wide conditions of CO_2 and CH_4 concentrations in CO_2/CH_4 mixture, total organic carbon (TOC), pressure, and temperature. Two different types of adsorption, namely excessive and absolute adsorption, are considered in the proposed models. Therefore, the model output is the adsorption value (excessive or absolute) of each CO_2 and CH_4 in the CO_2/CH_4 mixture and the input variables are CO_2 concentration, CH_4 concentration, TOC, temperature, pressure, type of adsorption (excessive and absolute), and type of component studied (CO_2 or CH_4). The developed models can estimate excessive and absolute adsorption values for CO_2 and CH_4 by using 0=excessive and 1=absolute, and 0=CO_2 and 1=CH_4 as input parameters for "type of adsorption" and "type of component studied", respectively. Excess adsorption refers to the volume of gas that is directly measured as being adsorbed in laboratory conditions. However, in reality, the presence of the adsorbed phase occupies some pore space, reducing the volume available for free gas. Excess adsorption does not account for this occupied volume, leading to an underestimation of the total adsorption capacity. Absolute adsorption, on the other hand, provides a more realistic representation of the actual adsorption occurring in the system. Therefore, determining these two types of adsorption is important.

Details of the used databank and operating conditions are shown in Table 1.

Table 1. Details and range of influencing factors used for designing the models construction

Parameter	Minimum	Maximum	Mean
Adsorption volume (cm^3/g)	0.004	12.5088	2.4693
CO_2 concentration (wt. %)	0	100	44.7562
CH_4 concentration (wt. %)	0	100	55.2438
TOC	1.05	5.14	2.8188
Temperature (°C)	35	85	45.9211
Pressure (MPa)	0.3422	11.2248	4.9698
Adsorption type		Excessive/Absolute	
Component type		CO_2/CH_4	

Shale samples used in this study have different characteristics in terms of mineralogical composition and TOC content. Eight shale samples were obtained from (Zhou et al., 2019), with seven being marine shales collected from the Silurian Longmaxi Formation (LMX1, LMX2, and LMX3) and the Ordovician Wufeng Formation (WF1, WF2, WF3, and XW) in the Sichuan Basin. The remaining sample is a continental shale taken from the Yanchang Formation (YC) in the Ordos Basin. WF3 has the highest content of quartz with 55.8 %, while XW has the highest amount of carbonate minerals with 35.8 %. The highest clay composition is related to the YC shale sample with 35.8 %. Continental shale generally exhibits higher clay content and lower quartz content in comparison to marine shale. Data of three shale samples were also gathered from (Luo et al., 2015). X-ray diffraction analysis of the shale samples showed that the mineral composition is primarily quartz, kaolinite, and illite, with smaller amounts of pyrite and K-feldspar. All of the studied samples contain significant quantities of clay minerals, making up to 70% in total. Kaolinite is the dominant clay mineral, accounting for over 50% by weight in most samples. Among them, the #3 shale sample exhibits higher clay mineral and kaolinite content compared to the other two samples. Data of CO_2 and CH_4 adsorption on four shale samples studied in (Gu et al., 2017) were also used in this work. The total organic carbon (TOC) content of the four shale samples ranges from 2.59 to 3.81 wt%. Since all values exceed 2 wt%, these shales are classified as organic-rich. The primary elements identified in the samples are silicon (Si), carbon (C), calcium (Ca), aluminum (Al), iron (Fe), potassium (K), magnesium (Mg), sodium (Na), and sulfur (S). Minor components such as titanium (Ti), sodium (Na), barium (Ba), and others are also present, but their oxide contents are significantly below 1 mass%. In addition, analysis on the shale samples used in the study of (Ma et al., 2019) revealed that the TOC contents are 3.03% and 5.40%. Additionally, the organic matter was found to be concentrated and classified as II 1- type. X-ray diffraction analysis confirmed that the primary mineral components of the shale include quartz, plagioclase, carbonates (such as calcite and dolomite), pyrite, and clay minerals. The brittle mineral content, including quartz, plagioclase, and carbonates, was measured at 2.75%, 3.9%, and 21.65%, respectively. On average, clay minerals constitute a significant portion, making up 40.75% of the composition. Moreover, data of four shale samples studied by (Qi et al., 2019) were utilized in this study. The TOC content across the samples is fairly consistent, varying from 2.28 to 3.61 wt%. Clay mineral content ranges between 15.7 and 27.4 wt%. The average pore sizes fall within the range of 4.72 to 5.56 nm, while the specific surface area of the samples varies from 13.86 to 19.12 m^2/g.

2.13. Model Development

In this study, a comprehensive databank was gathered from different references in the literature, and it was randomly divided into two subsets, training and testing subsets. For these subsets, 80% of total data ws utilized as train subset and the remaining 20% was considered as test subset. Accuracy and robutness of various adsorption models are evaluated using different statistical parameters including coefficient of determination (R^2), average absolute relative error (AARE %), root mean squared error (RMSE), and relative error (RE%). These parameters are defined in below equations:

$$AARE(\%) = \frac{\sum_{i=1}^{N} \frac{|Predicted_i - Experimental_i|}{Experimental_i}}{N} * 100 \tag{8}$$

$$R^2 = 1 - \frac{(Experimental_i - Predicted_i)^2}{(Experimental_i - mean\ of\ experimental\ data)^2} \tag{9}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Experimental_i - Predicted_i)^2}{N}} \tag{10}$$

$$RE(\%) = \frac{Predicted_i - Experimental_i}{Experimental_i} * 100 \tag{11}$$

In the above relations, N is the total number of data.

ANN models have the ability to correlate a nonlinear function with using only one hidden layer. In this work, one hidden layer was considered for designing the structure of ANN in order to reduce the computational time. A structure of 25-x-1 with twenty five neurons and one neuron in input and output layers, respectively was used for the proposed ANN.An input layer with 25 neurons (expanded from the 7 original input features), one hidden layer with 17 neurons, and an output layer with 1 neuron (25-17-1 structure). The tansig transfer function was used for the input and hidden layers, while the purelin transfer function was used for the output layer. Finally, the ANN model with the AARE % was used as the optimum structure of the developed ANN. Results of various ANNs with different number of neurons in the hidden layer (x) for train data are illustrated in Figure 6. As indicated in this plot, the network with x=17 (i.e. 25-17-1 structure) yields the least AARE (%) and is used for further studies.

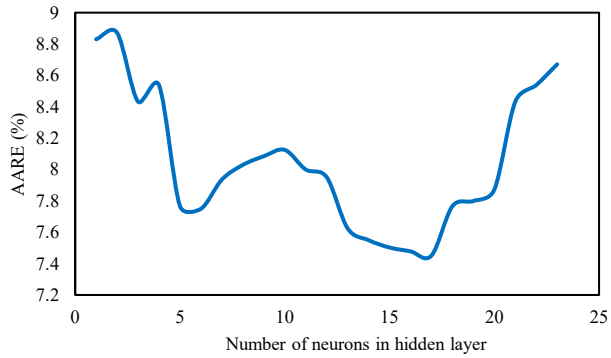


Figure 6. Performance of the ANN model under various numbers of neurons in the hidden layer

In order to design the structure of the LSSVM model, two factors, σ^2 and γ, need to be tuned. For this purpose, the coupled simulated annealing (CSA) algorithm was used. Based on the optimization process, the optimum values for σ^2 and γ were determined to be 8.975 and 111.09, respectively. For the adaptive boosting (AB) model, the number of base estimators as a tuning hyperparameter was adjusted based on the value of R^2. Figure 7 shows the output of AB model versus R^2 for different base estimators. According to this figure, the optimum values of 49 was obtained for the number of estimators, which led to the highest value of R^2.

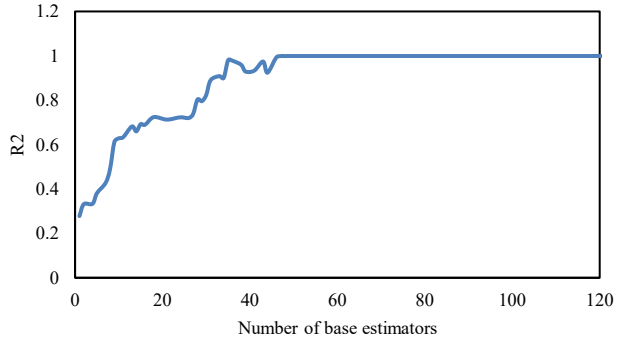


Figure 7. Coefficient of determination versus number of base estimators for the developed AB model

The maximum depth of DT model as a hyper-parameter controls its overall complexity. Therefore, it needs to be tuned for model over- and under-fitting prevention. The DT model results were evaluated based on the value of R^2 for the train subset, and a set of 1 to 10 was used for the maximum depth. As indicated in Figure 8, the optimum value of maximum depth is 6, as it yields the highest R^2 of 0.999.

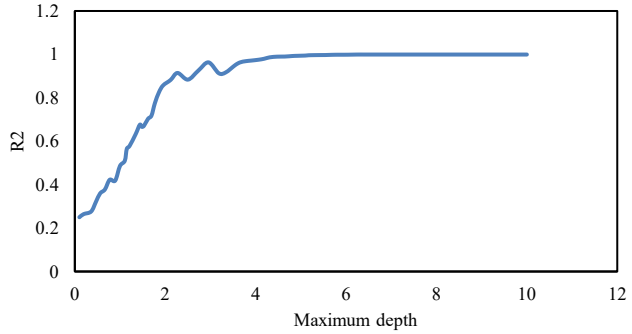


Figure 8. Performance of the developed DT model at different maximum depths

For the EL model, three approaches including LSSVM, KNN, and DT were used in parallel in the model structure. The RF model has two hyper-parameters of the number of estimators and maximum depth as uncertain parameters that need to be tuned during model construction. Based on the obtained R^2 value, the optimum maximum depth was estimated to be 8.2 and it yields the highest R^2 for all the numbers of estimators. For the KNN mode, the k value was found to be 8 as the developed model has the highest R^2 at this point, Figure 9.
The Performance of the developed models is then evaluated via statistical and graphical analyses, and their robustness in accurate estimation of CO_2 and CH_4 adsorption on shale rocks is confirmed by comparing their results with the results of well-established adsorption models including BET, D-R, and D-A.

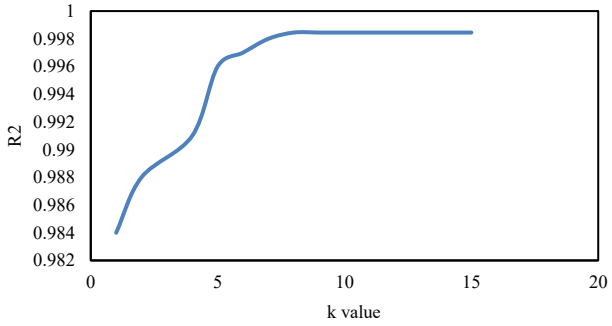
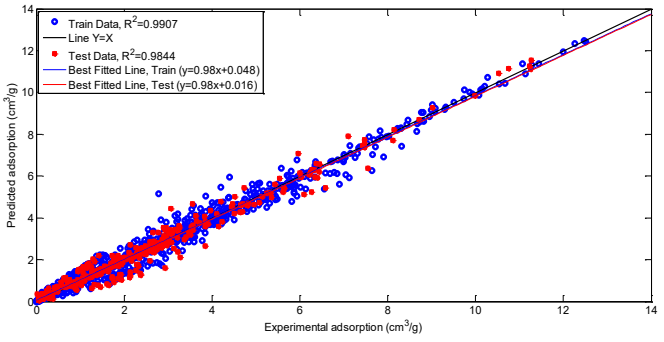


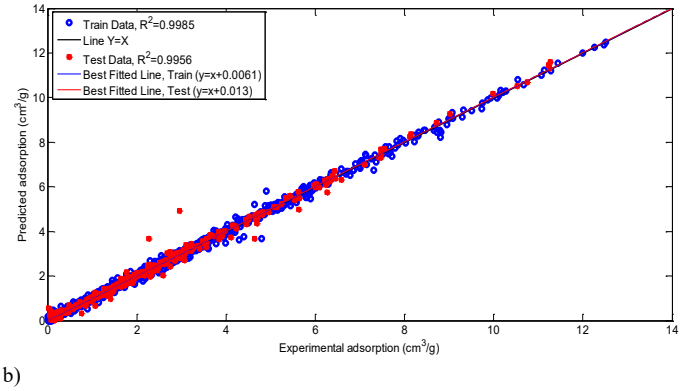
Figure 9. Effect of k value on the performance of the KNN model

3. Results and Discussion

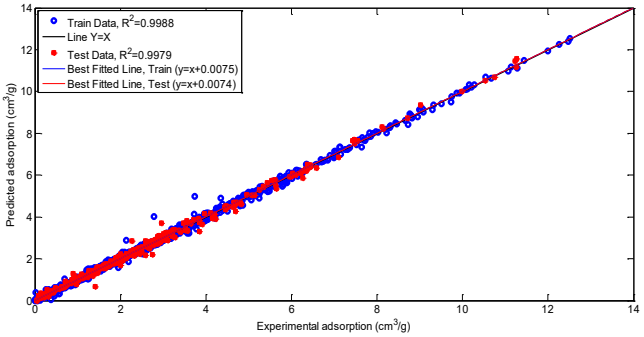
In this study, various data-driven models have been developed for accurate estimation of CO_2 and CH_4 adsorption on shale rocks in CO_2/CH_4 mixture. The models are applicable for wide ranges of pressure, temperature, TOC, CO_2 and CH_4 concentrations and are able to estimate both excessive and absolute adsorptions. Efficiency and accuracy of the developed models were evaluated through graphical and statistical analyses and the results of the proposed models were also compared with those of well-established adsorption models. Statistical parameters including R^2, AARE, RMSE, and RE were utilized for this purpose. In the first step, the model's accuracy in predicting CO_2 and CH_4 adsorption on shale rocks is shown through cross plots. The estimated data by different models for train and test subsets were plotted versus experimental values, Figure 10 (a-k). This figure indicates that the predicted adsorption values by the developed models for both train and test datasets are mostly concentrated around the line of y=x or the line with a slope of 45°, which shows the high accuracy of the presented models. Values of R^2 for different models are also shown in Table 2, for train, test, and total data. Based on this analysis, the random forest and decision tree models exhibit higher accuracy compared to the other data-driven models as well as popular adsorption models found in literature.



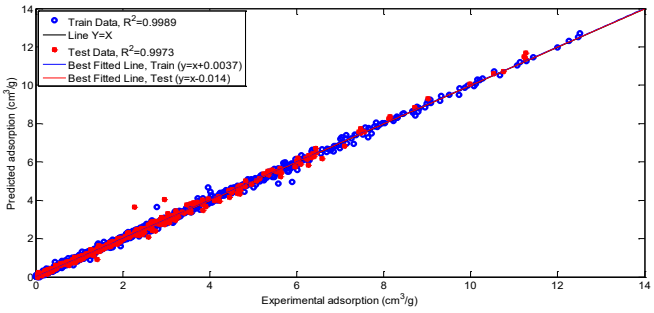
a)



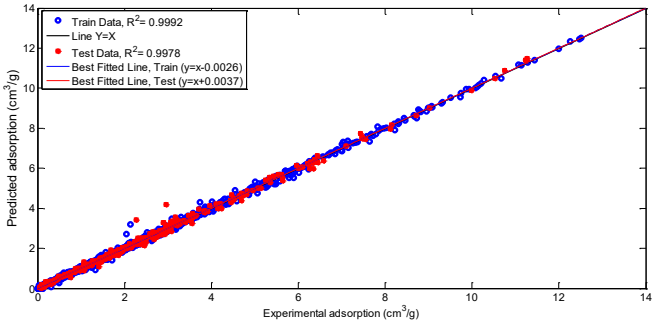
b)



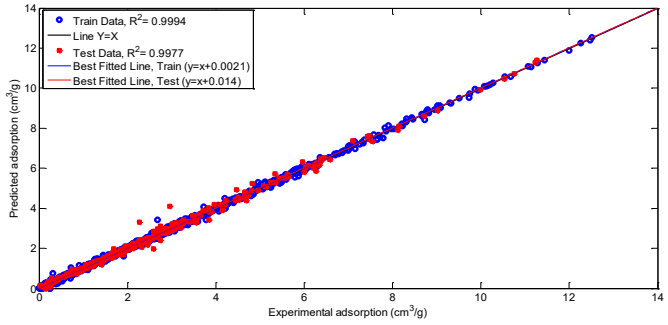
c)



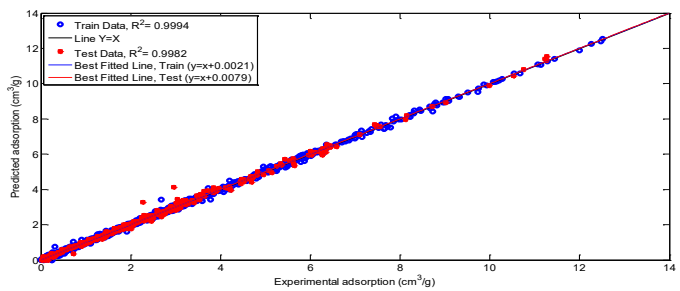
d)



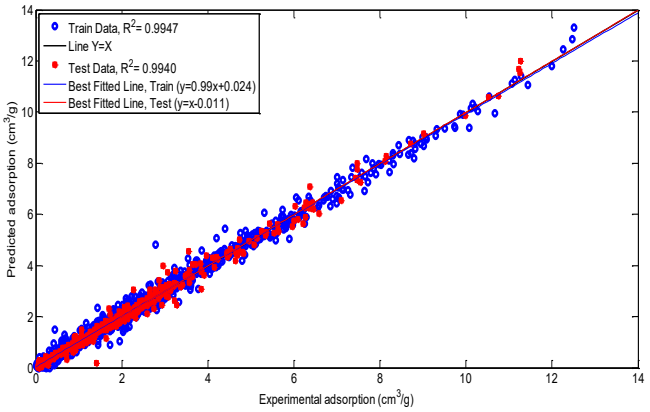
e)



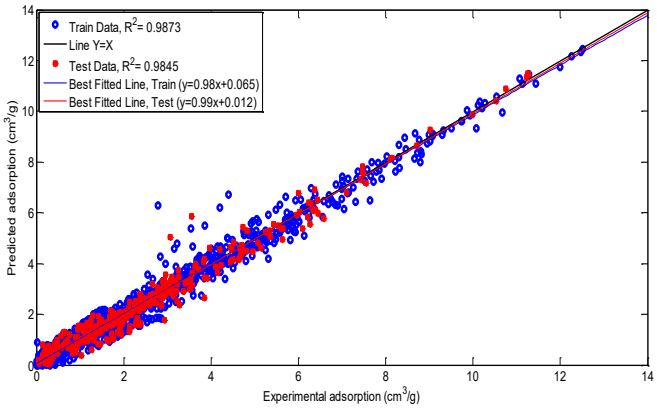
f)



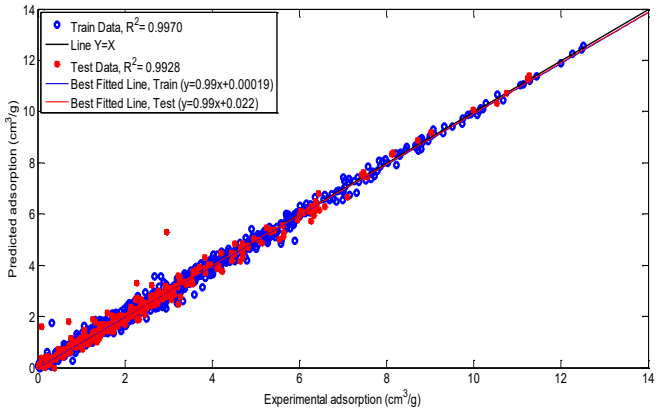
g)



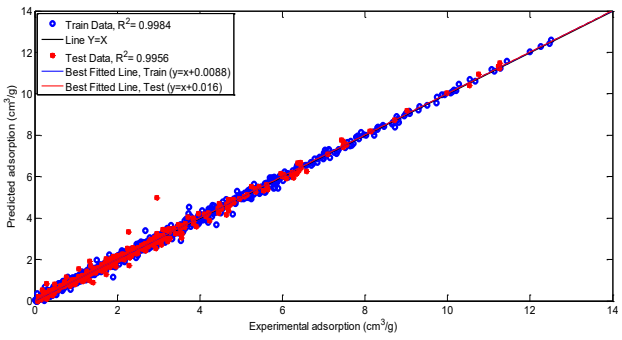
h)



i)

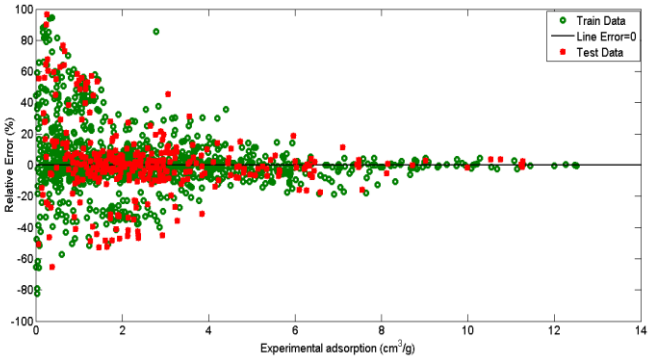


j)

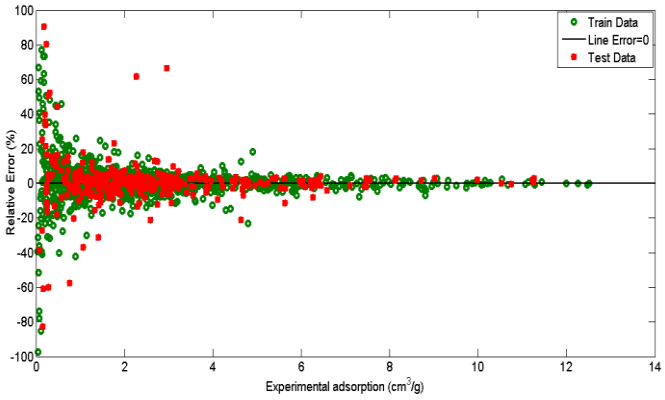


k)

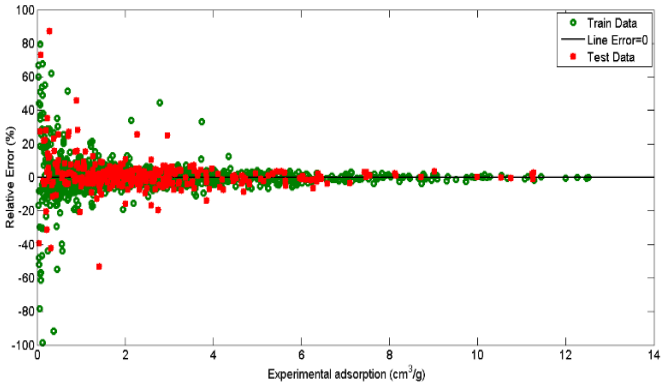
Figure 10. Cross plot of the predictions for different models; a) LSSVM, b) KNN, c) EL, d) ANN, e) AB, f) DT, g) RF, h) D-A, i) D-R, j) Langmuir, k) BET


Train Data
Line Error=0
Test Data
Relative Error (%)
Experimental adsorption (cm3/g)

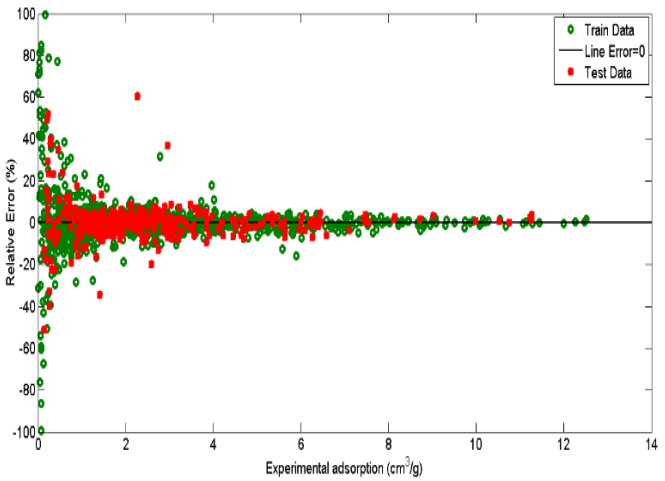
a)


Train Data
Line Error=0
Test Data
Relative Error (%)
Experimental adsorption (cm3/g)

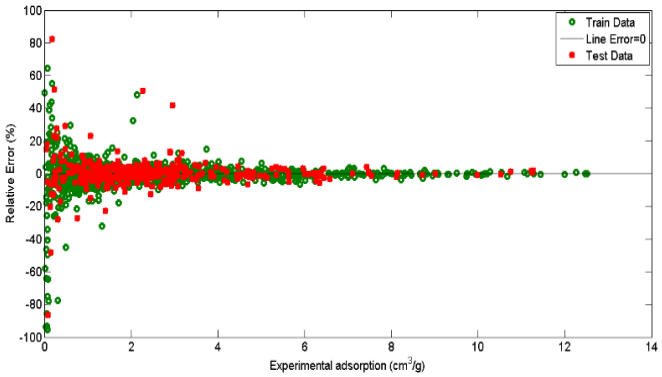
b)


Train Data
Line Error=0
Test Data
Relative Error (%)
Experimental adsorption (cm3/g)

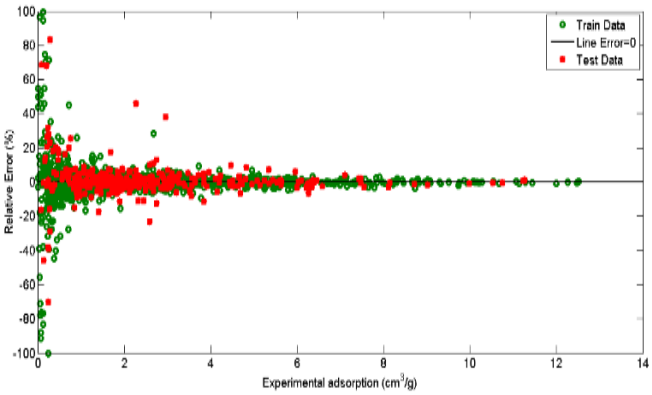
c)


Train Data
Line Error=0
Test Data
Relative Error (%)
Experimental adsorption (cm3/g)

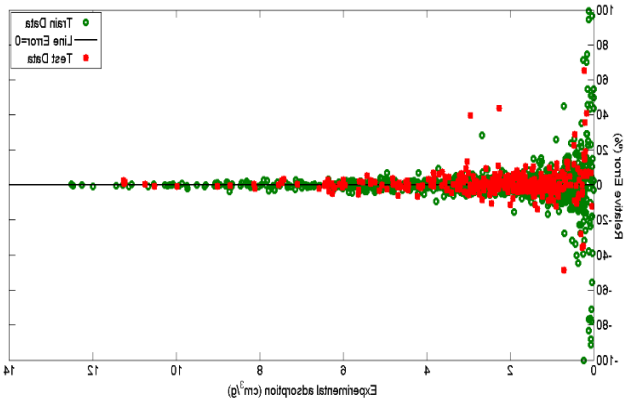
d)


Train Data
Line Error=0
Test Data
Relative Error (%)
Experimental adsorption (cm3/g)

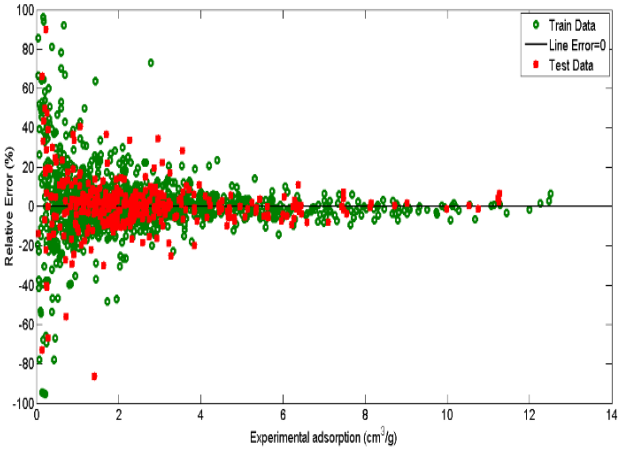
e)


Train Data
Line Error=0
Test Data
Relative Error (%)
Experimental adsorption (cm3/g)

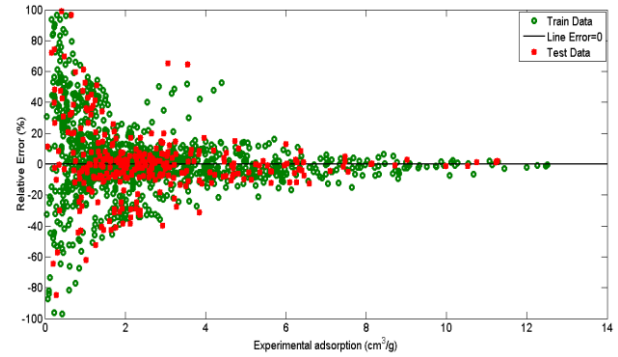
f)



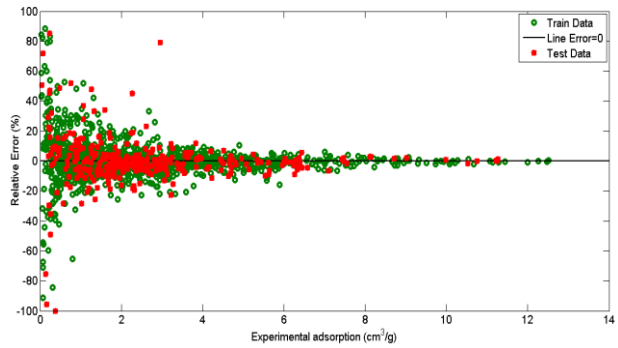
g)


Train Data
Line Error=0
Test Data
Relative Error (%)
Experimental adsorption (cm3/g)

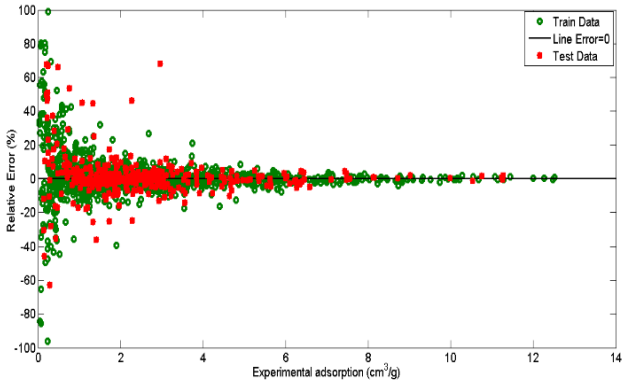
h)



i)



j)

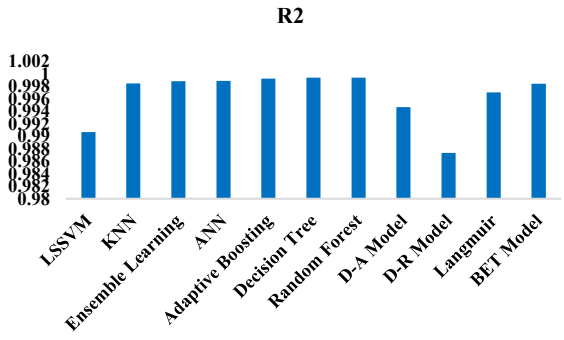


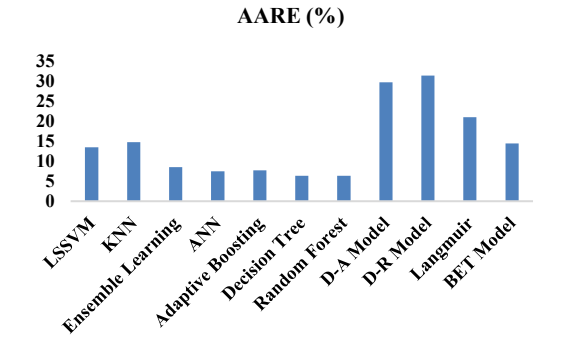
k)

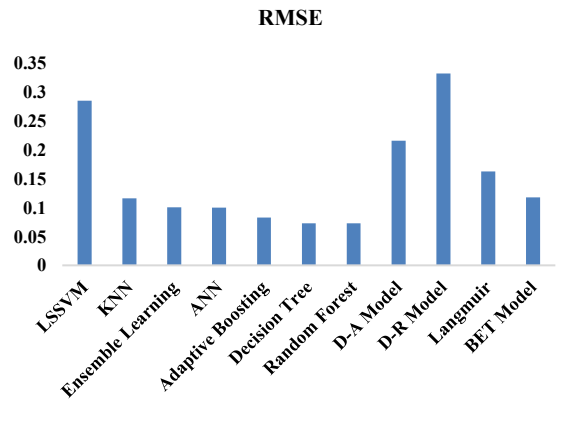
Figure 11. Relative error of the predictions for different models; a) LSSVM, b) KNN, c) EL, d) ANN, e) AB, f) DT, g) RF, h) D-A, i) D-R, j) Langmuir, k) BET

Table 2. Statistical analysis for different models in predicting CO2 and CH4 adsorption on shale rocks

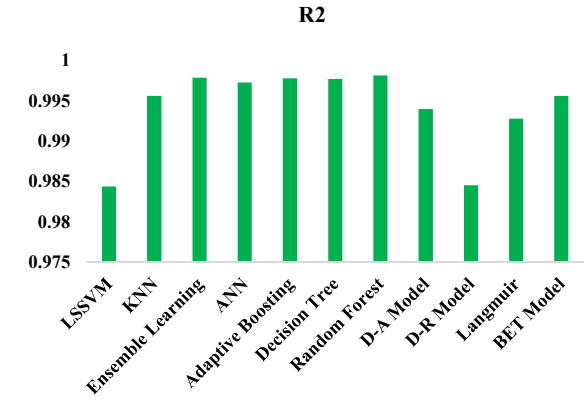
Models	R^2			AARE (%)			RMSE		
	Train	Test	Total	Train	Test	Total	Train	Test	Total
LSSVM	0.9907	0.9844	0.9894	13.4472	25.0915	15.7747	0.2856	0.3766	0.3059
KNN	0.9985	0.9956	0.9979	14.7147	24.0466	16.5926	0.1164	0.1999	0.1372
Ensemble Learning	0.9988	0.9979	0.9986	8.5374	10.0235	8.8344	0.1007	0.1388	0.1094
ANN	0.9989	0.9973	0.9985	7.4507	12.196	8.3993	0.1001	0.1578	0.1140
Adaptive Boosting	0.9992	0.9978	0.9989	7.6965	11.2166	8.4001	0.0830	0.1415	0.0975
Decision Tree	0.9994	0.9977	0.9991	6.3701	10.2122	7.1381	0.0729	0.1433	0.0914
Random Forest	0.9994	0.9982	0.9991	6.3711	8.0769	6.6304	0.0729	0.1307	0.0915
D-A	0.9947	0.9940	0.9945	29.6801	17.0206	27.1873	0.2161	0.2349	0.2200
D-R	0.9873	0.9845	0.9867	31.3377	31.0463	31.3189	0.3324	0.3751	0.3414
Langmuir	0.9970	0.9928	0.9962	20.9336	21.2564	21.0391	0.1625	0.2552	0.1848
BET	0.9984	0.9956	0.9978	14.3767	18.5786	15.2745	0.1180	0.1998	0.1383

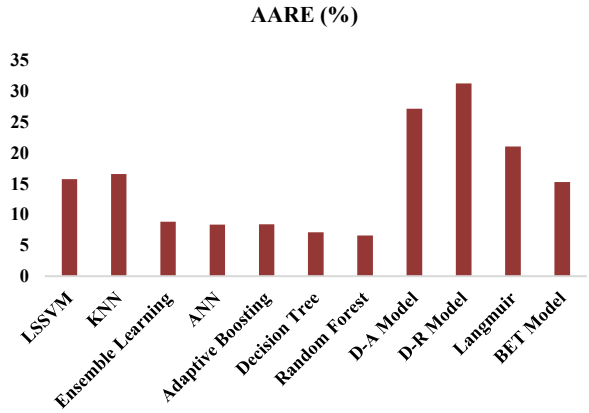


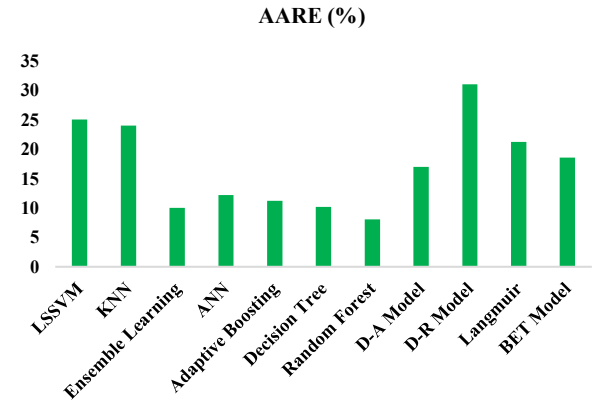


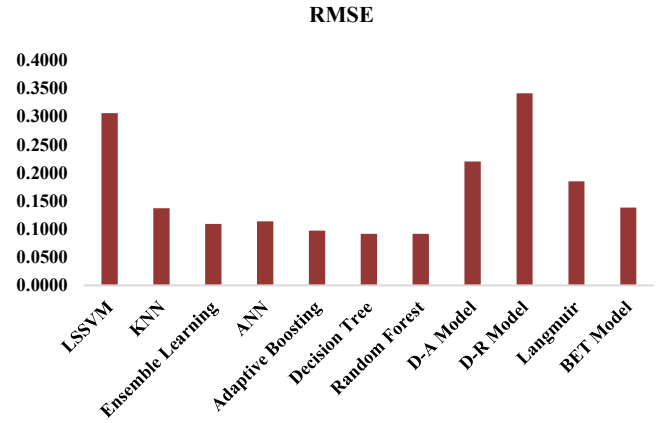


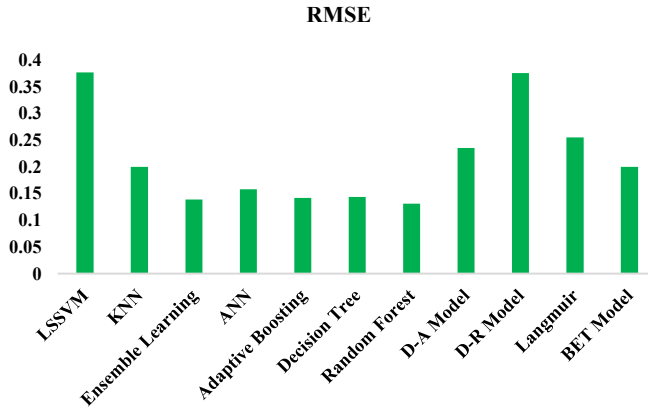
a)

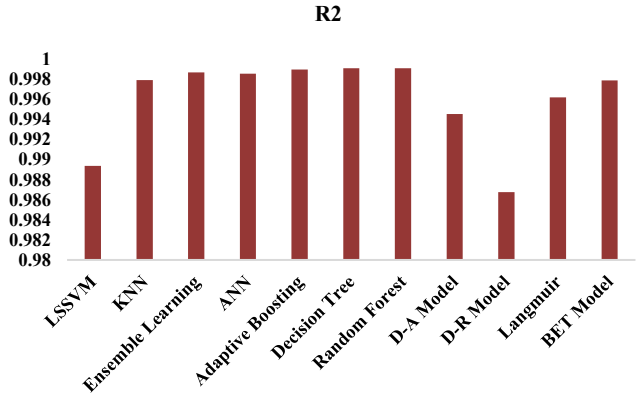












b)

Figure 12. Performance of different models in predicting CO2 and CH4 adsorption on shale rocks; a) Train data, b) Test data, c) Total data

According to the discussed statistical analyses, some developed models exhibit high accuracy and there is a great overlap between the experimental and predicted adsorption data of CO_2 and CH_4 on shale rocks. For further verification and in order to show the distribution of the obtained absolute relative errors by different models, data frequency is plotted against absolute relative errors, Figure 13. This figure shows that for decision tree and random forest models, about 90% of the total predicted adsorption values have errors less than 10%, while for D-R model, this error value is for 60% of total predicted outcomes. This analysis shows the robustness of the developed DT and RF models.

Predicting the adsorption capacity of different shale rocks and also determining the competitive adsorption of CO_2 and CH_4 on a shale sample is important in CO_2 geological storage as well as shale gas recovery [65]. Therefore, accurate modelling of CO_2 and CH_4 adsorption at various operating conditions is crucial. In this work, efforts have been performed to propose new efficient and robust models for simulating the adsorption behavior of CO_2 and CH_4 on shale rocks. According to the obtained results, the proposed EL, ANN, AB, DT, and RF models yield acceptable adsorption results, and among the developed models, DT and RF models have the highest accuracy, as shown in Figure 13. Moreover, it is worth mentioning that while models like ANN offered high accuracy, it requires longer training times compared to faster but slightly less accurate models like KNN and EL. In addition, all models were evaluated using k-fold cross-validation (k=10) to ensure robustness and generalization ability. The results confirm that the best-performing models maintain their accuracy across different subsets of the dataset.

Finding the effect of each influencing factor on the amount of adsorbed CO_2 and CH_4 is important in order to optimize the process of CO_2 storage in shale rocks or enhance shale gas recovery by CO_2 injection. Regarding this issue, sensitivity analysis was performed in this study using the Monte Carlo algorithm[66, 67]. In this algorithm, uncertainties and sensitivities are determined by using the results of multi-model variations with probabilistic model inputs[66, 67]. Results of sensitivity analysis for input parameters of CO_2 and CH_4 concentrations, pressure, temperature, and TOC of shale rock are shown in Figure 15. Sensitivity of each parameter is found by calculating changes in the output relative to the changes in the input parameter. As illustrated in Figure 14, pressure and TOC are the most effective parameters and can highly affect the amount of adsorbed CO_2 and CH_4 on a shale sample.

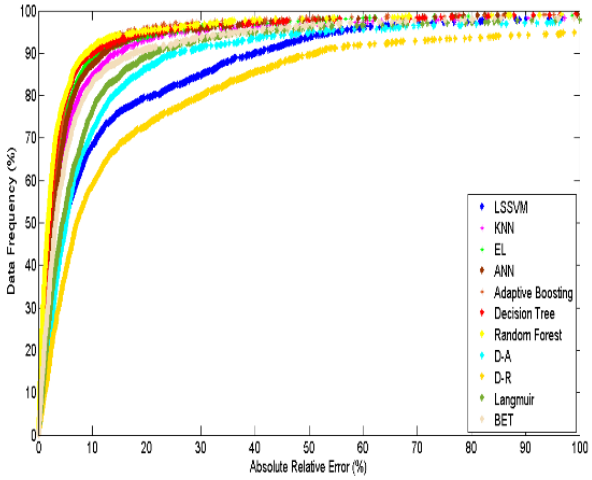


Figure 13. Cumulative data frequency versus absolute relative error of different models for total data

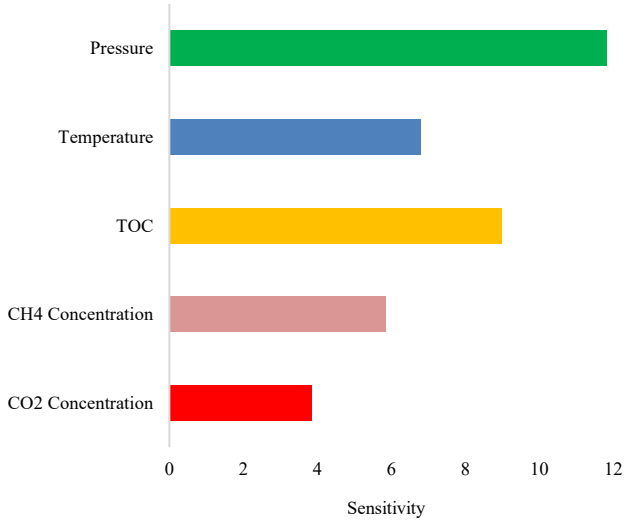


Figure 14. Sensitivity analysis for finding the importance of each influencing factor

6. Conclusions

In this study, seven intelligent, data-driven modeling approaches—Least Squares Support Vector Machine (LSSVM), K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), Decision Tree (DT), Adaptive Boosting (AdaBoost), Random Forest (RF), and Ensemble Learning—were developed and applied to model and predict both excess and absolute adsorption of CO_2 and CH_4 on shale rocks under varying operational conditions. The models incorporated key influencing parameters such as pressure, temperature, total organic carbon (TOC), and gas composition (CO_2/CH_4 concentrations).
The performance, accuracy, and robustness of the proposed models were rigorously evaluated using comprehensive statistical and graphical analyses. Furthermore, the predictive capabilities of these data-driven models were benchmarked against well-established theoretical adsorption models—including Langmuir, BET, Dubinin–Radushkevich (D-R), and Dubinin–Astakhov (D-A)—demonstrating superior accuracy and reliability in estimating adsorption behavior across diverse shale types. Among the developed models, the Decision Tree (DT) with R^2 of 0.9991 and AARE % of 7.138 and the Random Forest (RF) model with R^2 of 0.9991 and AARE % of 6.6304 exhibited the highest predictive performance, indicating their potential as powerful tools for simulating adsorption processes in complex geological media. Additionally, a sensitivity analysis conducted using the Monte Carlo simulation technique identified pressure and total organic carbon (TOC) as the most influential input variables affecting CO_2 and CH_4 adsorption capacity in shale rocks.

These models offer significant advantages in practical applications such as enhancing the accuracy of CO_2 and CH_4 adsorption predictions in shale formations, supporting the optimization of geological CO_2 sequestration strategies, improving recovery efficiency in shale gas reservoirs, and reducing reliance on costly and time-consuming experimental procedures. Overall, the proposed data-driven models contribute to a more efficient and accurate understanding of gas adsorption mechanisms in unconventional reservoirs, offering valuable insights for both scientific research and industrial applications.

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Nomenclature

TOC	Total Organic Carbon
ANN	Artificial Neural Network
SVM	Support Vector Machine
LSSVM	Least Square Support Vector Machine
DT	Decision Tree
BET	Brunauer-Emmett-Teller
D-A	Dubinin-Astakhov
D-R	Dubinin-Radushkevich
$ScCO_2$	Supercritical CO_2
BA	Bangham Adsorption
GCMC	Grand Canonical Monte Carlo
D-P	Dubinin-Polanyi
GMDH	Group Method Of Data Handling
GEP	Gene Expression Programming
PSO	Particle Swarm Optimization
MSE	Mean Square Error
RBF	Radial Basis Function
KNN	K-Nearest Neighbor
RF	Random Forest
Q	(mmol/g) CO_2 Adsorbed
Q_l	(mg/g) maximum monolayer coverage
k_l	Constant
Qm	Saturated Adsorption Capacity
P_0	Saturated Vapor Pressure
P	Equilibrium Pressure
D	Affinity Constant
CSA	Coupled Simulated Annealing
C	BET Constant
Q_0	Maximum Adsorption Volume
A	Structural Heterogeneity Constant of Shale
n_{op}	Number of Output Parameters
n_{dp}	Number of Data Patterns
t_{ij}	Target Values
y_{ij}	Network Outputs
α_i	Coefficient Associated With The Ith Feature
x_i	Support Vectors
$K(x,x_i)$	Vector Returned by The Chosen Kernel
B	Intercept of The Regression
C	Hyperparameters
E	SVR algorithm
Γ	SVR algorithm

Author Contributions

All authors contributed equally to the conception, preparation, and writing of this manuscript.

Conflict of Interest

The authors, while observing publication ethics in the referencing, declare the absence of interest of conflict.

Acknowledgement

The authors gratefully acknowledges the support provided by the Petroleum University of Technology (PUT) of in conducting this research.

AI-Assisted

The authors retain full responsibility for the accuracy, originality, integrity, legality, and ethical compliance of all text, data, tables, figures, analyses, references, and conclusions, and have verified all AI-assisted outputs. Any fabrication, falsification, manipulation, plagiarism, copyright infringement, or other misuse of AI remains solely the responsibility of the authors and does not transfer or diminish their accountability for the published work.

References

[1] A. Daryasafar, A. Keykhosravi, K. Shahbazi. *Journal of Cleaner Production*, 2019, **239**, 1181011-14.
[2] A. A. Tanha, et al. *Petroleum Research*, 2023, **8(1)**, 63-70.
[3] F. A. Hameli, et al. *Energy Science & Engineering*, 2024.
[4] L. Hou, et al. *Renewable and Sustainable Energy Reviews*, 2024, **189**, 1138781-9.
[5] A. A. Tanha, et al. *Petroleum & Coal*, 2022, **64(2)**, 25-38.
[6] S. Zhang, et al. *Fuel*, 2025, **398**, 1355911-13.
[7] Q. Li, et al. *Natural Resources Research*, 2025, **34(3)**, 16671-14.
[8] Q. Li, et al. *Processes*, 2025, **13(6)**, 16301-9.

[9] Q. Liao, et al. *Fuel*, 2023, **339**, 1274001-13.
[10] Z. Lv, et al. *Applied Energy*, 2024, **355**, 1222421-18.
[11] D. Mani, et al. *Shale Gas*, 2017, 25-54.
[12] R. Heller, M. Zoback. *Journal of Unconventional Oil and Gas Resources*, 2014, **8**, 14-24.
[13] S. Merey, C. Sinayuc. *International Journal of Oil, Gas and Coal Technology*, 2018, **17(2)**, 172-188.
[14] X. Du, et al. *Fuel*, 2021, **283**, 1188861-13.
[15] C. Qin, et al. *Energy*, 2020, **206**, 1181501-8.
[16] J. Zhou, et al. *Fuel*, 2019, **251**, 293-306.
[17] Z. Li, et al. *Chemical Engineering Journal*, 2022, **428**, 1325261-12.
[18] T. F. Rexer, et al. *Energy & Fuels*, 2013, **27(6)**, 3099-3109.
[19] C. R. Clarkson, B. Haghshenas. *SPE Unconventional Resources Conference/Gas Technology Symposium*, 2013.
[20] A. Fatah, et al. *Energies*, 2020, **13(12)**, 32001-13.
[21] B. X. Medina-Rodriguez, V. Alvarado. *Energies*, 2021, **14(10)**, 28801-16.
[22] X. Tang, et al. *Energy & Fuels*, 2017, **31(10)**, 10787-10801.
[23] M. Aly, A. Abdulraheem, T. Moawad. *SPE Canadian Energy Technology Conference*, 2023.
[24] M. Huang, et al. *Chemical Engineering Journal*, 2023, **474**, 1455621-14.
[25] M. Meng, R. Zhong, Z. Wei. *Fuel*, 2020, **278**, 1183581-12.
[26] H. Wang, et al. *Applied Energy*, 2024, **370**, 1236381-9.
[27] L. Wang, et al. *SPE Europec Featured at EAGE Conference and Exhibition*, 2020.
[28] H. Ma, et al. *SPE Western Regional Meeting*, 2023.
[29] M. N. Amar, et al. *Journal of Petroleum Science and Engineering*, 2022, **208**, 1092261-22.
[30] A. Bemani, et al. *Journal of Natural Gas Science and Engineering*, 2020, **76**, 1032041-14.\
[31] A. Daryasafar, A. Ahadi, R. Kharrat. *The Scientific World Journal*, 2014, **2014(1)**, 2465891-10.
[32] M. Rahmati, G. Zargar, A. A. Tanha. *Petroleum Research*, 2024, **9(2)**, 176-192.
[33] P. Vaziri, B. Sedaee. *Clean Energy*, 2024, **8(1)**, 111-125.
[34] N. Sipöcz, F. A. Tobiesen, M. Assadi. *Applied Energy*, 2011, **88(7)**, 2368-2376.
[35] A. Daryasafar, K. Shahbazi. *Petroleum Science and Technology*, 2018, **36(5)**, 333-337.
[36] A. Baghban, et al. *International Journal of Greenhouse Gas Control*, 2017, **57**, 143-161.
[37] A. Baghban, A. H. Mohammadi, M. S. Taleghani. *International Journal of Greenhouse Gas Control*, 2017, **58**, 19-41.
[38] X. Wang, C. W. Chan, T. Li. *Chemical Engineering Science*, 2024, **290**, 1198781-15.
[39] U. Di Caprio, et al. *Journal of CO2 Utilization*, 2023, **70**, 1024521-12.
[40] M. Hosseinpour, et al. *Fuel*, 2023, **353**, 1292651-13.
[41] L. Feng, J. Zhu, Z. Jiang. *Journal of Cleaner Production*, 2024, **450**, 1419161-17.
[42] A. Alanazi, et al. *International Journal of Coal Geology*, 2023, **275**, 1042971-14.
[43] X. Gong, et al. *Atmosphere*, 2024, **15(1)**, 118-127.
[44] S. P. Mousavi, et al. *Scientific Reports*, 2023, **13(1)**, 79461-24.
[45] H. Bahmaninia, et al. *Fluid Phase Equilibria*, 2023, **567**, 1137121-15.
[46] R. Soleimani, A. H. S. Dehaghani. *Journal of Environmental Management*, 2024, **354**, 1202981-12.
[47] D. M. Makarov, et al. *Separation and Purification Technology*, 2023, **325**, 1246141-20.
[48] B. Liu, et al. *Separation and Purification Technology*, 2024, 129306, **20(8)**,1-18.
[49] A. Mahmoudzadeh, et al. *Korean Journal of Chemical Engineering*, 2024, **41(5)**, 1427-1448.
[50] S. E. Jerng, Y. J. Park, J. Li. *Energy and AI*, 2024, 1003611-14.
[51] S. Pakseresht, M. Kazemeini, M. M. Akbarnejad. *Separation and Purification Technology*, 2002, **28(1)**, 53-60.
[52] S. I. Hansen, et al. *The Journal of Physical Chemistry C*, 2025, **129(11)**, 5393-5407.
[53] A. Bahadori, H. B. Vuthaluru. *International Journal of Greenhouse Gas Control*, 2009, **3(6)**, 768-772.
[54] S. Brunauer, P. H. Emmett, E. Teller. *Journal of the American Chemical Society*, 1938, **60(2)**, 309-319.
[55] M. Mahmoud, et al. *Journal of Petroleum Science and Engineering*, 2020, **188**, 1069231-10.
[56] J. Villarroel-Rocha, et al. *Journal of Porous Materials*, 2024, **31(3)**, 1111-1120.
[57] K. H. Chu, et al. *Industrial & Engineering Chemistry Research*, 2024.
[58] Z. Hussain, et al. *Energy & Fuels*, 2026, **40(16)**, 8997-9016.
[59] I. Ushiki, Y. Tahara. *Journal of CO2 Utilization*, 2024, **84**, 1028571-13.
[60] M. Gu, et al. *Journal of Natural Gas Science and Engineering*, 2017, **46**, 296-306.
[61] X. Luo, et al. *International Journal of Coal Geology*, 2015, **150**, 210-223.
[62] Y. Ma, et al. *Carbon Resources Conversion*, 2019, **2(1)**, 35-42.
[63] R. Qi, et al. *Fuel*, 2019, **242**, 728-743.
[64] H. Zhao, T. Wu, A. Firoozabadi. *Fuel*, 2018, **224**, 412-423.
[65] A. Daryasafar, et al. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 2020, 1-18.
[66] J. C. Helton. *Reliability Engineering & System Safety*, 1993, **42(2-3)**, 327-367.
[67] R. L. Iman, J. C. Helton, J. E. Campbell. *Journal of Quality Technology*, 1981, **13(4)**, 232-240.