Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predictive Analysis of CO2 and CH4 Sorption Mechanisms in Unconventional Reservoirs

Abu Bakker Siddique

abubakker.siddiqu@sus.edu

Minhaz Chowdhury

Shahjalal University of Science and Technology

Mahamudul Hashan

Shahjalal University of Science and Technology

Labiba Nusrat Jahan

Shahjalal University of Science and Technology

Md. Iftiar Uddin

Shahjalal University of Science and Technology

Research Article

Keywords: CH_4 sorption, CO_2 sorption, Total organic carbon, Machine learning, Unconventional reservoir

Posted Date: March 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-7473704/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Predictive Analysis of CO_2 and CH_4 Sorption Mechanisms in Unconventional Reservoirs

Abu Bakker Siddique*[1], Minhaz Chowdhury[2], Mahamudul Hashan[2], Labiba Nusrat Jahan[2], Md. Iftiar Uddin[2]

[1] Department of Mechanical Engineering, Southern University and A&M College, Baton Rouge, LA 70807

[2]Department of Petroleum and Mining Engineering, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh

*Corresponding Author: E-mail: abubakker.siddiqu@sus.edu; Tel.: +12253713933

Abstract

This study presents a robust machine learning framework for accurately predicting the sorption capacities of two greenhouse gases, CH_4 and CO_2, in unconventional reservoirs under diverse thermodynamic conditions. A comprehensive set of ensemble and hybrid models including ExtraTrees, XGBoost, LightGBM, Gradient Boosting, CatBoost, and four newly developed hybrid architectures (HM1–HM4) was trained and evaluated using an extensive experimental dataset. The hybrid models demonstrated consistently superior performance, with the best CH_4 model (HM3) achieving an R^2 of 0.9905 and the best CO_2 model (HM3) reaching an R^2 of 0.9949 on unseen test data. Unlike previous studies, where only selected models performed well for either CH_4 or CO_2, the proposed models exhibited high accuracy and generalizability across both gas types. Model interpretability was enhanced through SHAP and Partial Dependence Plot (PDP) analyses. For CH_4 sorption, total organic carbon (TOC) and pressure were identified as the most influential features, while for CO_2 sorption, Gas (%) representing CO_2 concentration was the dominant factor, followed by temperature and pressure. PDP analysis further revealed a strong linear relationship between CO_2 sorption and Gas (%), with TOC contributing significantly at lower levels before plateauing. Moisture content was found to have a mild negative effect on sorption in both cases. These results confirm the physical consistency of the models and their capacity to capture complex sorption behavior, offering valuable tools for gas-in-place estimation, CO_2 storage evaluation, and production forecasting in unconventional reservoirs.

Keywords: CH_4 sorption, CO_2 sorption, Total organic carbon, Machine learning, Unconventional reservoir.

Article Highlights

- Hybrid Model 3 achieved the highest R^2 for CH_4 (0.9905) and CO_2 (0.9949) sorption prediction.
- SHAP and PDP analyses identified TOC as key feature for CH_4; Gas (%) for CO_2.
- Models captured the decline in CO_2 sorption at high pressures, consistent with adsorption behavior.
- All models were validated under varying thermodynamic and geological conditions.

1. Introduction

Adsorption is a key mechanism in purification and separation processes (Beutekamp & Harting, 2002; King, 2013), valued in environmental chemistry for its cost-effectiveness and efficiency in removing contaminants across broad concentration ranges without complex infrastructure (Dong et al., 2010; Rebouh et al., 2016). Gas adsorption is particularly relevant for methane (CH_4) and carbon dioxide (CO_2) sequestration in subsurface formations (Goetz et al., 2006). Unconventional resources such as coalbed methane (CBM) and shale formations show significant greenhouse gas storage potential (Chalmers & Bustin, 2008; Ross & Bustin, 2007; Yang et al., 2011; Yang & Liu, 2020).

Shale gas storage involves both surface adsorption and free gas in pores, contrasting with CBM reservoirs dominated by adsorption and conventional reservoirs containing mainly free gas (Martin et al., 1998; Hill et al., 2007; Bernard et al., 2010; Glorioso & Rattia, 2012). CO_2 injection in enhanced CBM (ECBM) and enhanced shale gas recovery (ESGR) displaces CH_4 due to the higher CO_2 affinity of coal and shale (Mukherjee & Mishra, 2018). Sorption occurs via chemisorption, involving strong chemical bonds, or physisorption, governed by van der Waals forces (Busch & Gensterbum, 2011). In shales, physisorption ranges from monolayer adsorption at low pressure to multilayer adsorption at high pressure (Kang et al., 2011). CO_2, despite its larger kinetic diameter, penetrates smaller pores more effectively and interacts with more functional groups than nonpolar CH_4, leading to higher CO_2 affinity (Zhang et al., 2016; Liu, 2020).

Reservoir adsorption capacity depends on pressure, temperature, rock composition, and injected gas properties (Crosdale et al., 1998). Total organic carbon (TOC) is a dominant factor (Lu et al., 1995; Chalmers & Bustin, 2006, 2008; Ross & Bustin, 2007, 2008; Weniger et al., 2010; Zhang et al., 2012; Gasparik et al., 2013), while moisture reduces sorption by competing for adsorption sites (Pan et al., 2010; Gensterbum et al., 2013; Rother et al., 2012; Gasparik et al., 2014; Shabani et al., 2018; Arunachalam, 2021). Sorption also increases with aromaticity and decreases with thermal maturity, while clay minerals especially montmorillonite can match or exceed TOC-rich shales in capacity (Gasparik et al., 2012; Ji et al., 2012).

CO_2-to-CH_4 selectivity (SCO_2/CH_4) rises with TOC and falls with clay content due to fewer high-energy adsorption sites in illite and kaolinite (Cao et al., 2025). Montmorillonite's expandable interlayer and cation exchange properties enhance CO_2 sorption, as shown in Grand Canonical Monte Carlo simulations (Hui et al., 2023). CO_2 generally adsorbs

faster than CH_4, though CH_4 may dominate at low pressures (Busch et al., 2003). Sorption hysteresis, particularly in CO_2 systems, may improve storage stability (Wang et al., 2014; Zhang et al., 2022).

Recent studies have increasingly applied machine learning and advanced simulation techniques to enhance CO_2 storage assessment, monitoring, and optimization in various geological settings (Hema et al., 2024; Iskandarov et al., 2024; Narayan et al., 2024; Li et al., 2025; Vishal et al., 2025). Both supervised and unsupervised ML algorithms have shown effectiveness in the field of adsorption (Bhattacharya, 2021; Okoroafor et al., 2022; Samnioti & Gagnais, 2023). Notably, ML models have been used to predict CO_2 sorption with high accuracy. For example, Rostami et al. (2018) develop an artificial neural network (ANN) model for CO_2 adsorption on activated carbon, achieving superior predictive performance compared to traditional isotherm models. Barki et al. (2019) apply ANN models on a large dataset covering single, binary, and ternary gas sorption scenarios, training dedicated models for each subset to improve prediction accuracy. Zhou (2019) demonstrates that the Group Method of Data Handling (GMDH) provides robust performance in predicting CO_2 adsorption on various commercial activated carbons, identifying the Brunauer–Emmett–Teller (BET) surface area as the most influential input parameter. Similarly, Bemani et al. (2020) employ the Least Squares Support Vector Machine (LSSVM) for sorption analysis, achieving reliable predictive outcomes. Huang et al. (2022) develop a fully connected neural network trained on molecular dynamics simulation data to model CH_4 sorption profiles in shale reservoirs, showcasing the utility of combining simulation-driven data with machine learning techniques. More recently, Alanazi et al. (2023) report promising results using machine learning models to predict CO_2 adsorption across various coal types and under different operating conditions. Tavakolian et al. (2023) compile a comprehensive dataset consisting of 3,804 data points of CH_4 and CO_2 sorption under diverse thermodynamic scenarios. Their model incorporates critical variables such as CO_2 injection percentage, rock type, total organic carbon (TOC), moisture content, temperature, and pressure. Leveraging Optuna for hyperparameter optimization, they identify the Random Forest algorithm as the most effective model, capable of accurately predicting both single and binary gas sorption behaviors. Notably, the model captures the atypical decline in CO_2 uptake at elevated pressures. The model achieves strong performance metrics, with mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) reported as follows: for CH_4, MAE = 0.0864, MSE = 0.0231, RMSE = 0.1520; and for CO_2, MAE = 0.0529, MSE = 0.0533, RMSE = 0.2308. Furthermore, Alqahtani et al. (2024) proposed a GWO-optimized CatBoost model, alongside GRNN and RBFNN, to predict CH_4 and CO_2 sorption in tight reservoirs using over 3800 data points. The CatBoost-GWO model achieved

outstanding accuracy (R^2 = 0.9993 for CO_2, 0.9970 for CH_4; RMSE = 0.1229 and 0.0681, respectively), ensuring both predictive power and physical consistency.

This study compiles a comprehensive set of experimental sorption data from tight reservoirs to develop robust predictive models for CH_4 and CO_2 sorption under varying thermodynamic and lithological conditions. A range of machine learning algorithms was employed, including ExtraTrees, XGBoost, LightGBM, Gradient Boosting, CatBoost, and four hybrid models, with ExtraTrees, LightGBM, and the application of four hybrid architectures and Partial Dependence Plots (PDP) introduced for the first time in this context. Hyperparameter tuning via GridSearchCV ensured optimal model performance. Across all models, features such as total organic carbon (TOC), CO2 concentration (Gas %), Pressure (MPa), Temperature(C), Moisture (%), and rock type were sufficient to accurately predict sorption behavior. SHAP and PDP analyses identified TOC as the most influential feature for CH_4 sorption, while CO_2 percentage and temperature were dominant for CO_2. Unlike previous studies, where only selected models performed well for either CH_4 or CO_2, the proposed models exhibited high accuracy and generalizability across both gas types. These findings highlight the improved generalization and reliability of the proposed models, positioning them as superior to those previously reported in the literature. Figure 1 illustrates the methodological workflow adopted in this study.

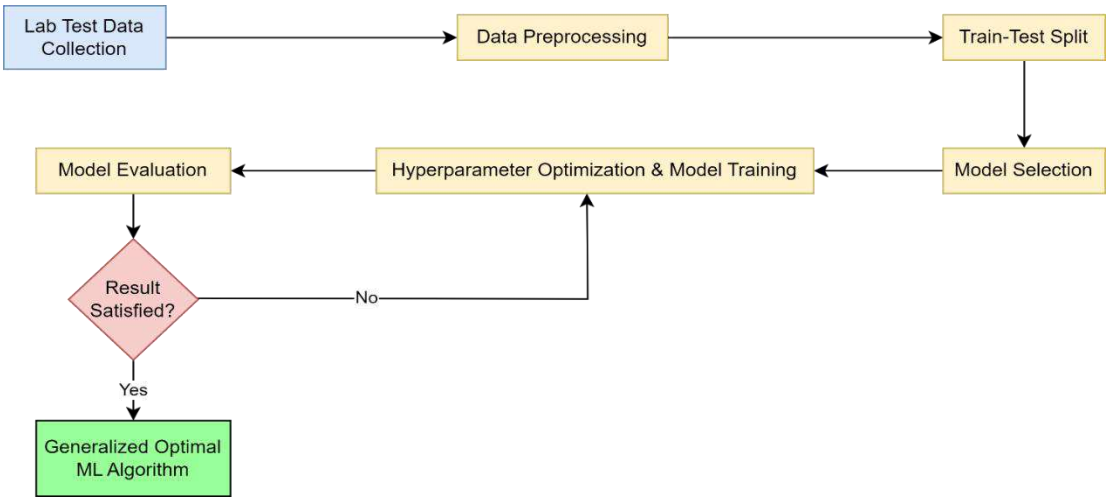


Figure 1: Workflow of the current study.

2. Theoretical Background

2.1 ExtraTrees (Extremely Randomized Trees)

ExtraTrees is an ensemble learning algorithm based on decision trees that introduces greater randomness during model construction to enhance its generalization ability. Unlike traditional decision trees, which attempt to find the optimal

split at each node, ExtraTrees selects random splits, reducing the chance of overfitting. It builds a collection of trees by selecting random subsets of features and data, leading to diverse predictions. The algorithm then aggregates these predictions to produce a final result. This randomness makes ExtraTrees faster and more robust than regular decision trees, particularly when dealing with high-dimensional data (Geurts et al., 2006). By averaging the predictions from multiple trees, ExtraTrees typically performs well in terms of accuracy while mitigating the risk of overfitting.

2.2 LightGBM (Light Gradient Boosting Machine)

LightGBM is an advanced gradient boosting framework optimized for processing large-scale datasets with high-dimensional features. It constructs decision trees in a sequential manner, with each tree aiming to minimize the residual errors of its predecessor. A key innovation of LightGBM lies in its histogram-based approach, where continuous feature values are grouped into discrete bins, significantly lowering computational overhead. Moreover, it incorporates Gradient-based One-Side Sampling (GOSS), a strategy that prioritizes the most informative instances for training, thereby reducing the effective dataset size while maintaining model accuracy. Renowned for its speed and scalability, LightGBM is particularly well-suited for real-time prediction scenarios (Ke et al., 2017). Another practical advantage is its native support for categorical features, eliminating the need for manual encoding.

2.3 CatBoost (Categorical Boosting)

CatBoost is a gradient boosting algorithm developed to improve the performance of machine learning models with categorical features. Unlike other boosting methods like XGBoost and LightGBM, which require preprocessing of categorical features (such as one-hot encoding), CatBoost handles them natively, making it particularly efficient for datasets with a large number of categorical variables. It employs a novel technique called ordered boosting, which addresses the overfitting problem that commonly occurs in traditional boosting algorithms by training trees based on different permutations of the data. This reduces bias in model predictions. Moreover, CatBoost uses target encoding for categorical variables, where the category values are replaced by the average target value for that category, which improves model accuracy without introducing data leakage (Prokhorenkova et al., 2018). These advancements make CatBoost particularly suitable for real-world datasets in industries like finance, healthcare, and e-commerce.

2.4 Gradient Boosting

Gradient Boosting is a powerful ensemble learning technique that builds predictive models in a stage-wise fashion, optimizing for accuracy by reducing prediction errors sequentially. Unlike bagging methods (like Random Forests and ExtraTrees) that create independent trees, Gradient Boosting builds each new tree to correct the residual errors of the previously built ensemble. At every step, the algorithm fits a new tree to the negative gradient of the loss function, effectively minimizing the model's error. The method is highly flexible and can be adapted to regression, classification, and ranking tasks using different loss functions. However, due to its sequential nature, Gradient Boosting is more prone to overfitting compared to bagging methods. To control this, techniques like learning rate, tree depth limitation, and subsampling are commonly used. Despite its sensitivity, when properly tuned, Gradient Boosting often delivers state-of-the-art performance in structured data problems and competitions (Friedman, 2001).

2.5 XGBoost (Extreme Gradient Boosting)

XGBoost is a widely used gradient boosting algorithm known for its exceptional performance in machine learning competitions. It builds an ensemble of decision trees, where each tree attempts to correct the errors made by the previous one, using a gradient descent technique. A key feature of XGBoost is its use of regularization, which penalizes overly complex models to prevent overfitting. It also employs a second-order approximation of the loss function, making it more computationally efficient than traditional gradient boosting methods. XGBoost is highly scalable and can be parallelized across multiple cores, making it well-suited for large-scale data. Its robust handling of sparse data and missing values adds to its popularity in practical machine learning tasks, from classification to regression (Chen & Guestrin, 2016). This combination of regularization, speed, and efficiency explains why XGBoost has become one of the most dominant models in machine learning.

2.6 Stacking Regressor

Stacked generalization, often referred to as stacking, is an ensemble learning strategy aimed at enhancing predictive performance by integrating the outputs of multiple base-level models. Instead of relying on a single model, stacking leverages the diverse learning capacities of various algorithms to collectively model complex data patterns. Initially, several base models are independently trained on the same dataset, and their respective predictions are then passed as inputs to a higher-level model commonly termed the meta-learner which generates the final prediction. This approach is particularly effective in mitigating the limitations inherent in individual models by capitalizing on their

complementary strengths. Stacking has been widely employed across different machine learning tasks, including classification, regression, and outlier detection, owing to its consistent improvements in accuracy and robustness

The concept of stacked generalization was first formalized by Wolpert in his seminal 1992 publication, where he introduced the theoretical foundation for blending multiple predictive models to reduce generalization error (Wolpert, 1992). This foundational work was further advanced in the comprehensive textbook The Elements of Statistical Learning, authored by Hastie, Tibshirani, and Friedman, which elaborated on ensemble techniques and their applications in real-world data problems (Hastie et al., 2009). Additionally, Breiman contributed significantly to the development of stacking methodologies through his influential study on "Stacked Regressions", where he demonstrated the effectiveness of combining regressors to improve model performance (Breiman, 1996). Together, these contributions have established stacking as a vital and reliable technique in the modern machine learning landscape.

3. Data collection and analysis

The dataset employed in this study consists of 3,805 data points compiled by Tavakolian et al. (2023), encompassing sorption measurements for CH_4, CO_2, and their mixtures under a range of thermodynamic conditions. It includes detailed information on the composition of the injected gases, rock types (shale/clay and coal/activated carbon), and several key input variables: CO2 Percentage (Gas ((%)), total organic carbon (TOC, %), temperature (°C), moisture content (%), and pressure (MPa). These parameters are used to predict the sorption capacities of methane (CH_4) and carbon dioxide (CO_2), expressed in cm^3/g. This comprehensive dataset provides a robust basis for model development, allowing for a nuanced investigation of sorption behavior across diverse geological and environmental settings.

Figure 2 presents the Pearson correlation heatmap illustrating the pairwise relationships among the key variables influencing gas presence and sorption behavior. A strong positive correlation is observe d between gas content and Total Organic Carbon (TOC) ($r = 0.61$), indicating that formations with higher organic richness tend to retain more gas. Similarly, gas content is strongly correlated with CO_2 excess sorption ($r = 0.58$), suggesting that formations capable of adsorbing CO_2 also tend to be gas-bearing likely due to shared dependencies on microporosity or organic matter. TOC is moderately correlated with CO_2 sorption ($r = 0.34$), reaffirming the role of organic components in enhancing adsorption. In contrast, CH_4 excess sorption shows a moderate negative correlation with both gas content ($r = –0.32$) and TOC ($r = –0.26$), possibly reflecting measurement conditions, adsorption competition, or structural heterogeneities in organic-rich samples. Moisture content shows weak correlations across all variables, with slight positive association

with TOC (r = 0.20) and negligible influence on gas (r = 0.05) or CH_4 sorption (r = –0.19). Temperature and pressure are moderately correlated (r = 0.26), consistent with expected geothermal gradients, but show weak or no correlations with sorption or gas presence, implying that pressure-temperature effects alone do not drive adsorption trends in this dataset. Overall, these results emphasize the importance of TOC and CO_2 sorption capacity as dominant indicators for gas-bearing formations, while CH_4 sorption exhibits complex or inverse relationships that warrant further investigation.

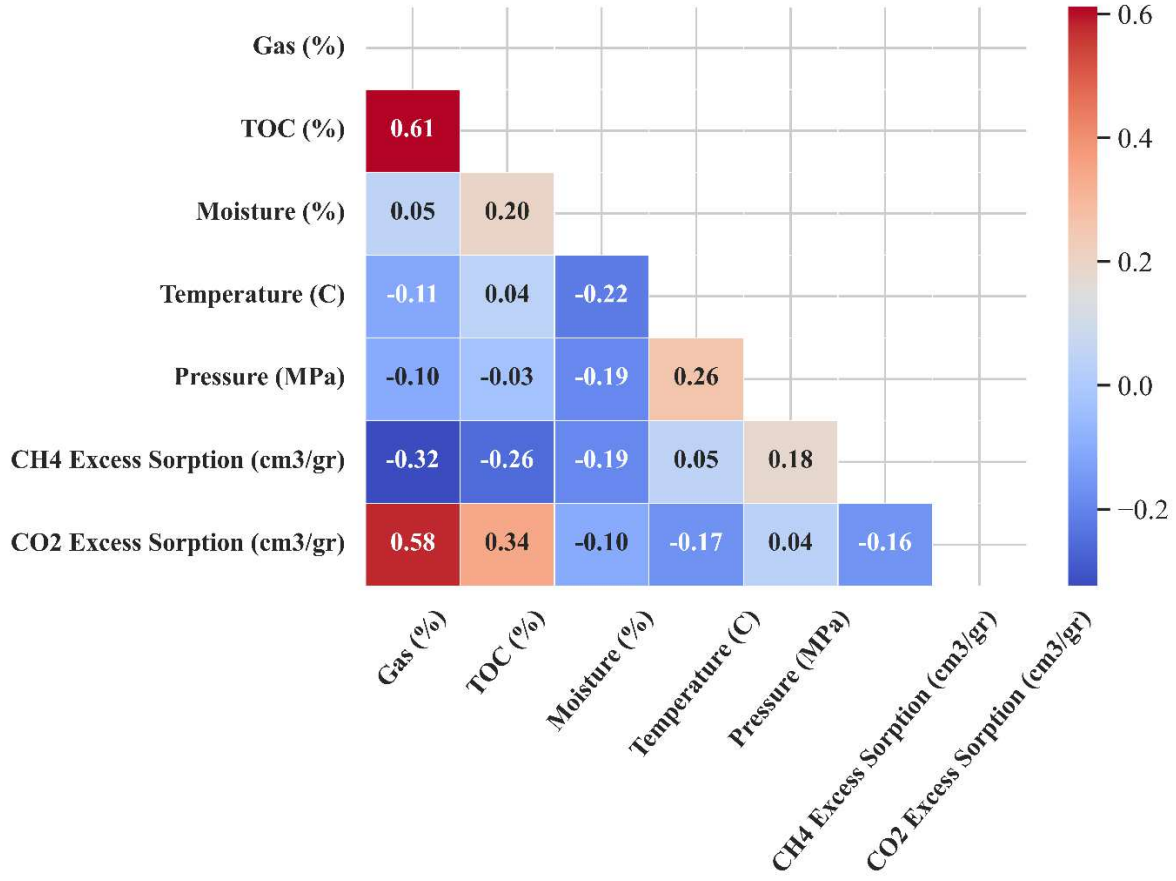


Figure 2: Correlation heatmap.

Table 1: Statistical summary.

	count	*mean*	*std*	*min*	*25%*	*50%*	*75%*	*Max*
Gas	3805	0.123	0.316	0	0	0	0	1
TOC	*3805*	*16.294*	*26.684*	*0.09*	*1.91*	*5.15*	*10.94*	*88.5*
Moisture	*3805*	*0.974*	*1.8*	*0*	*0*	*0*	*1.34*	*10.97*
Temperature	*3805*	*57.038*	*23.588*	*25*	*45*	*50.4*	*65*	*150*
Pressure	*3805*	*11.313*	*7.565*	*0*	*5.25424*	*9.94*	*17.23*	*29.247*
CH4 Excess Sorption	*3805*	*1.2402*	*1.2623*	*0*	*0.31*	*0.31*	*1.71*	*7.0362*
CO2 Excess Sorption	*3805*	*0.53*	*2.261*	*0*	*0*	*0*	*0*	*24.468*

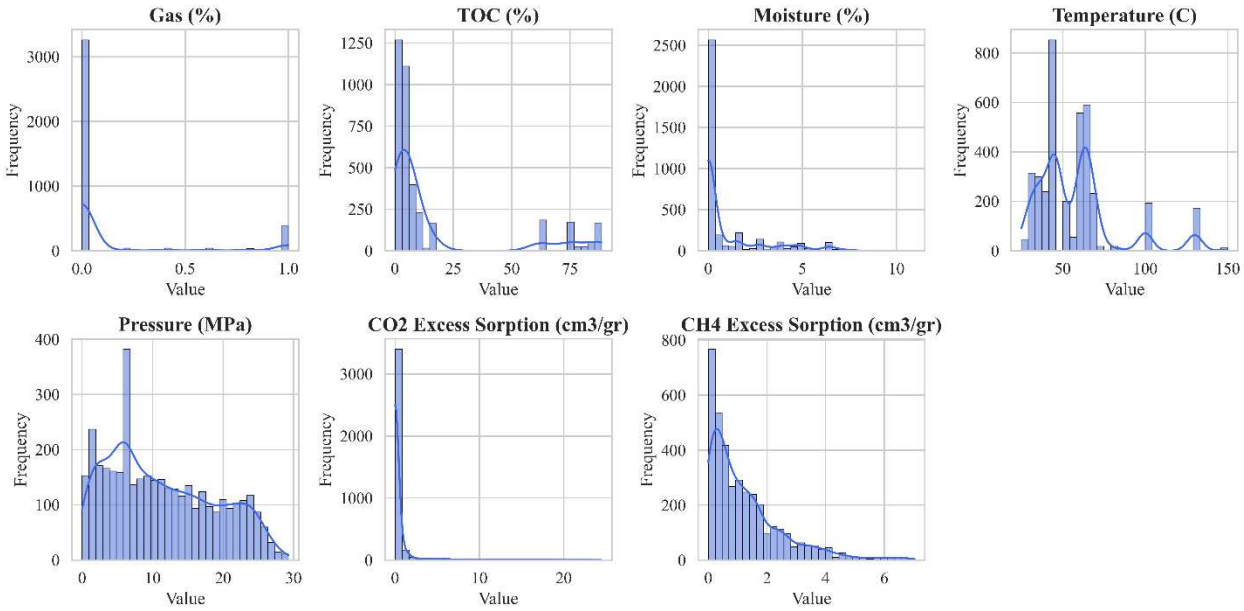


Figure 3: Histogram of input and output features

Table 1 presents the statistical summary of key input and output variables used in this study, while Figure 3 illustrates their distribution through histograms. The Gas (%) variable exhibits a mean of 0.123 and a standard deviation of 0.316, with only 12.3% of the samples showing gas presence. This indicates a highly imbalanced distribution skewed toward zero, posing challenges for predictive modeling. Total Organic Carbon (TOC) spans a broad range (0.09–88.5%) and exhibits a high standard deviation (26.684%), reflecting strong right-skewness due to a predominance of low-TOC samples and the presence of a few highly organic-rich formations. Moisture content follows a similar pattern, with a median of 0% and a mean of 0.974%, suggesting that most samples are nearly dry, punctuated by occasional outliers with higher water content. The temperature distribution, centered around 57 °C, and pressure, with a mean of 11.31 MPa, show relatively more symmetric shapes but still exhibit multimodal and broad variability, indicative of diverse

subsurface thermodynamic environments. CH_4 excess sorption shows a right-skewed, continuous distribution, concentrated at lower values (median = 0.31 cm^3/g) with a long tail representing higher adsorption capacities in select samples. CO_2 excess sorption, on the other hand, is sharply peaked near zero, with over half the samples showing negligible adsorption and only a limited number reaching values up to 24.47 cm^3/g, reflecting a highly sparse sorption landscape. These distributional characteristics, visualized in Figure 3, underscore the non-Gaussian and heterogeneous nature of the dataset, which is critical to account for in downstream modeling. Appropriate data preprocessingsuch as normalization, outlier handling, and balancing techniques is thus essential to ensure robust model performance across the diverse geological and geochemical profiles captured in this study. Figure 4 shows the pair plot of input and output features.

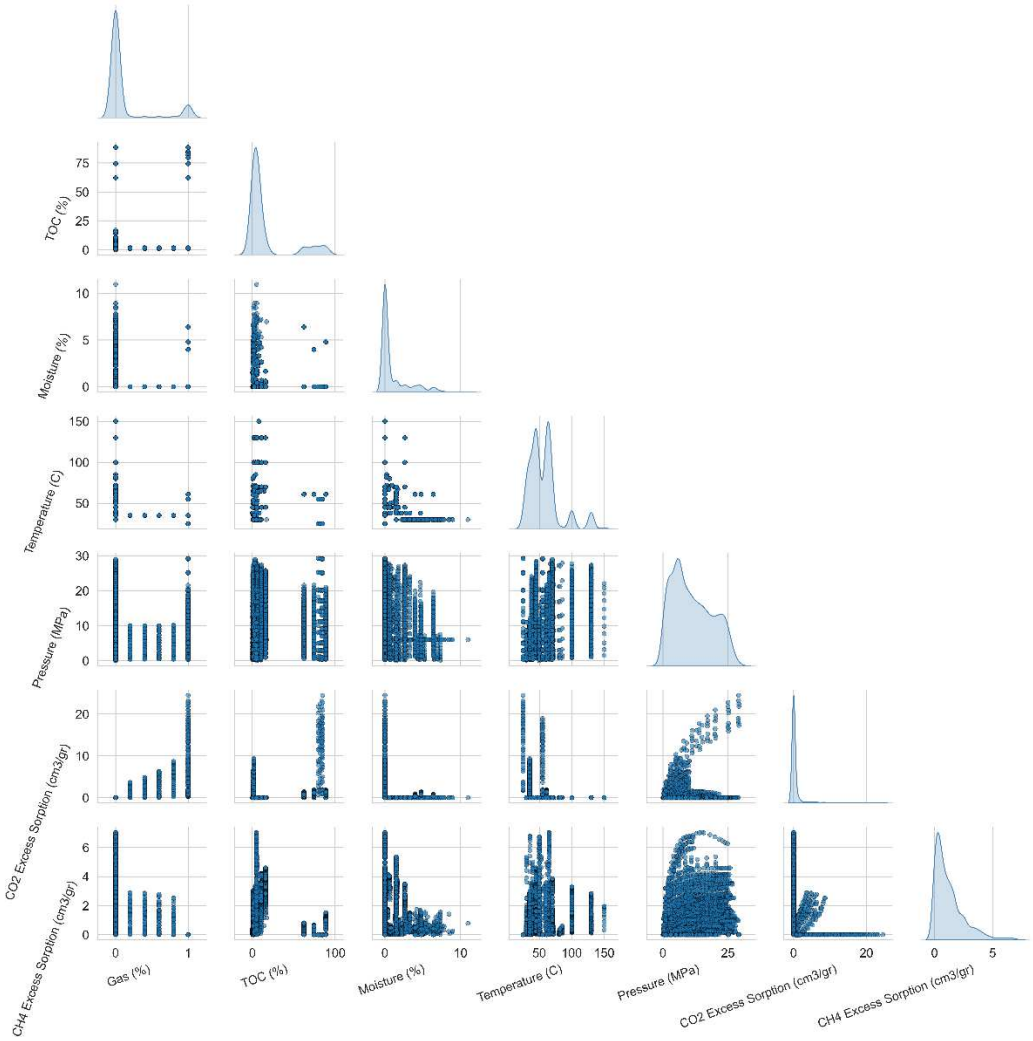


Figure 4: Pairplot of input and output features

Figure 5 and Figure 6 present the relationships between excess sorption of CO_2 and CH_4 and four key parameters: temperature, total organic carbon (TOC), moisture content, and pressure. These relationships are central to understanding gas storage and transport in subsurface formations, particularly for applications such as CO_2 sequestration and shale gas production. Below is the the description of the Figure 5 and Figure 6:

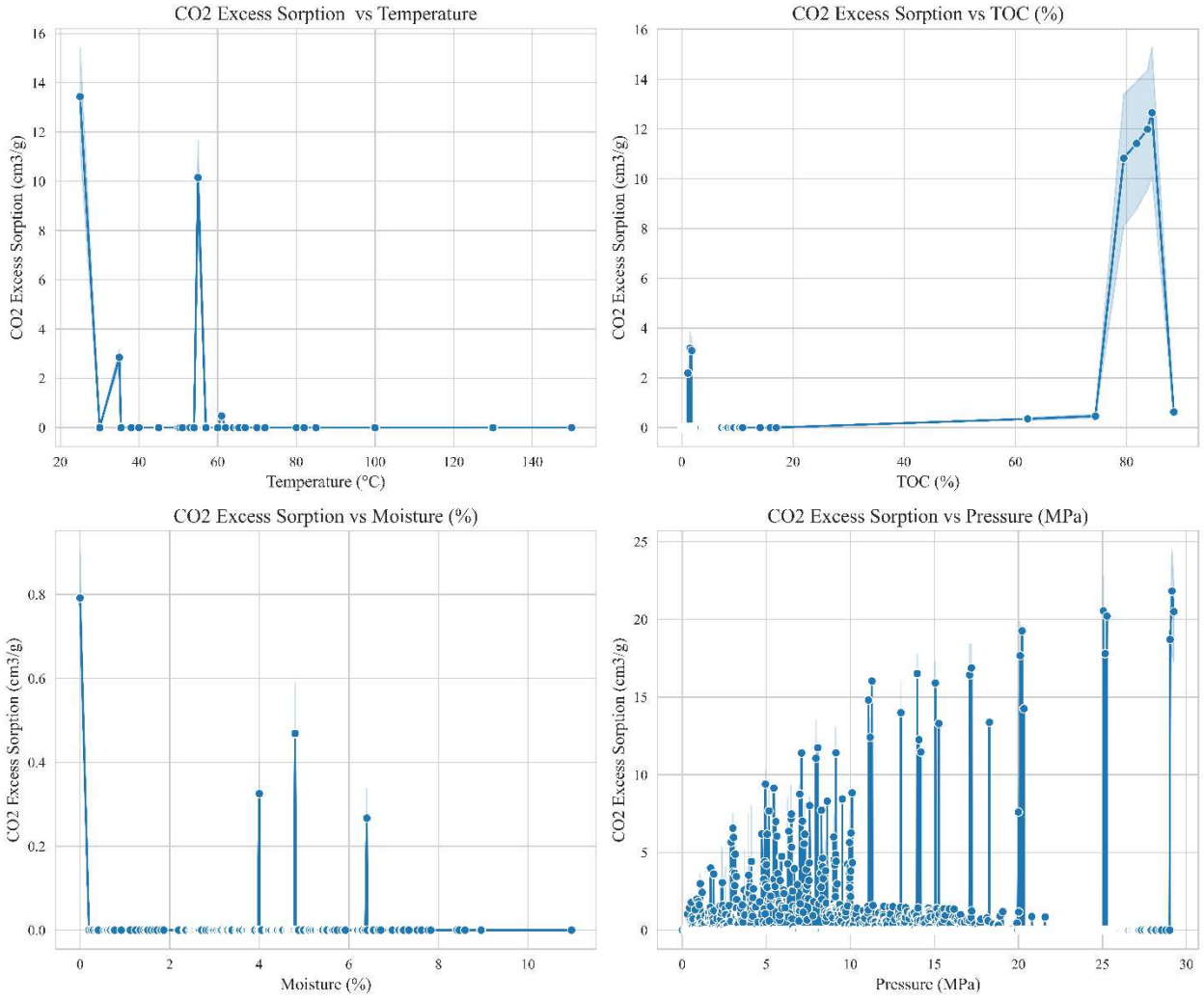


Figure 5: CO2 Sorption Profile Respect of inputs

- **Excess Sorption vs Temperature**

1. CO_2: The plot shows a sharp decline in CO_2 excess sorption with increasing temperature, with the highest sorption at lower temperatures (20–40 °C), and near-zero sorption above 60 °C.

2. CH_4: CH_4 sorption displays a more complex, less monotonic trend, but generally, sorption decreases with increasing temperature, especially above 80 °C.

- **Excess Sorption vs TOC (%)**

1. CO_2: Sorption remains low at TOC < 60%, but increases sharply at very high TOC (> 80%).
2. CH_4: Sorption is highest at low TOC and decreases as TOC increases.

- **Excess Sorption vs Moisture (%)**

1. CO_2 & CH_4: Both gases show a marked decrease in excess sorption with increasing moisture content, with sorption approaching zero above ~4–6% moisture.

- **Excess Sorption vs Pressure (MPa)**

1. CO_2: Excess sorption increases with pressure, showing a rapid rise at low pressures and a tendency toward saturation or a plateau at higher pressures.
2. CH_4: Sorption also increases with pressure, but the plateau is reached at lower sorption values compared to CO_2.

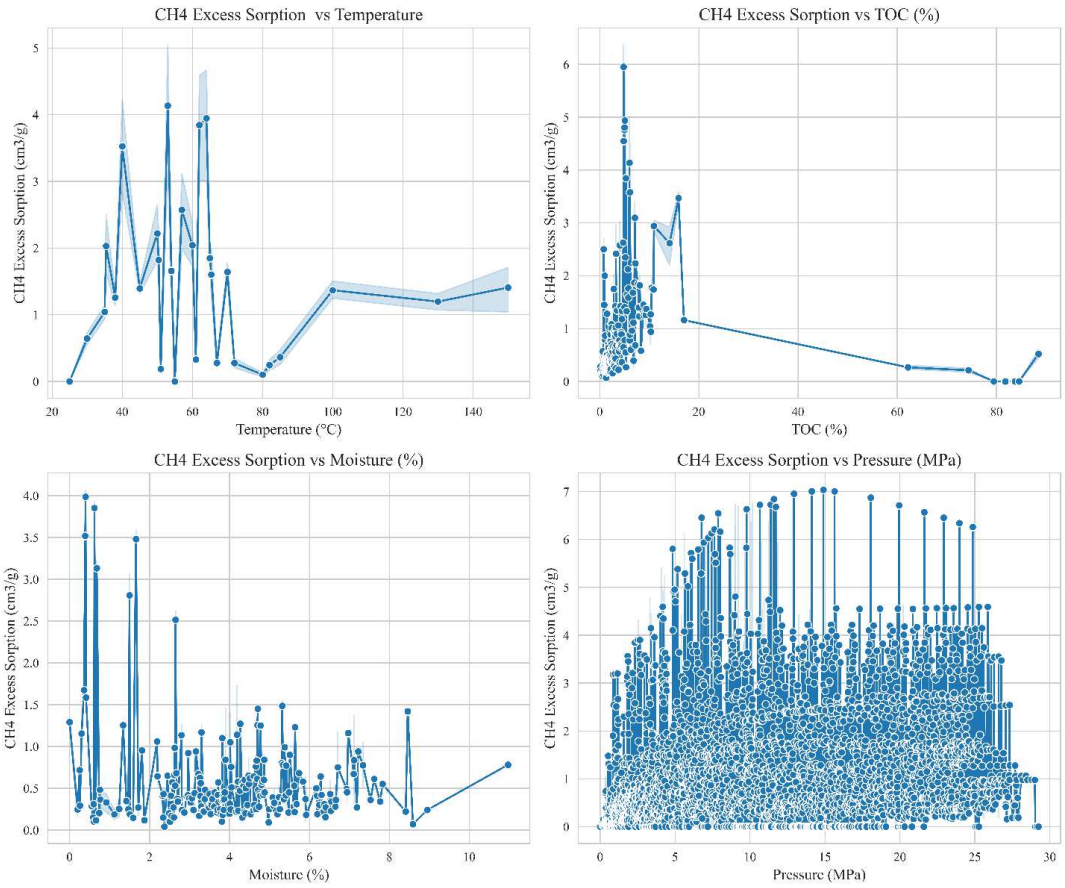


Figure 6: CH4 Sorption Profile Respect of inputs

4. Methodology

4.1. Data Partitioning Strategy

The entire dataset was randomly divided into two subsets: 80% for training and 20% for testing. This stratified approach ensured that the model was developed using a substantial portion of the data while maintaining a separate testing set to objectively assess its generalization capability on previously unseen instances.

4.2. Cross-Validation and Hyperparameter Optimization

To enhance the model's generalization ability and reduce the likelihood of overfitting, Grid Search Cross-Validation (GridSearchCV) was utilized alongside a 5-fold cross-validation framework. The training dataset was partitioned into five equal subsets, with the model trained on four subsets and validated on the remaining one in each iteration. This procedure was systematically repeated across all folds and combinations of hyperparameters to identify the optimal configuration. The optimal hyperparameters derived from this process are summarized in Table 2.

4.3. Ensemble Learning Models and Tuning Framework

To accurately predict the sorption capacities of CO_2 and CH_4, five state-of-the-art ensemble regression algorithms were employed: ExtraTrees, LightGBM, CatBoost, Gradient Boosting, and XGBoost. Each model underwent exhaustive hyperparameter optimization via grid search to enhance prediction accuracy and minimize error metrics. The optimization strategies and selected hyperparameters for each gas type is presented in Table 2.

Table 2: Hyperparameters of models.

Models	Hyperparameters utilized	Best Parameters CO2	Best Parameters CH4
ExtraTrees	"n_estimators": [200,300,400,500], "max_depth": [5,8,10,12],	max_depth=12, n_estimators=200	max_depth=12, n_estimators=400
LightGBM	"n_estimators": [200,300,400,500], "max_depth": [5,8,10,12], "learning_rate": [0.001, 0.01,],	learning_rate=0.01, max_depth=12, n_estimators=500	learning_rate=0.01, max_depth=12, n_estimators=500
CatBoost	"iterations": [200,300,400,500],"max_depth": [5,8,10,12], "learning_rate": [0.001, 0.01,],	Iterations=500, Learning_rate=0.01, Ma x_depth=12	Iterations=500, Learning_rate=0.01, max_depth=12
GradientBoosting	"n_estimators": [200,300,400,500], "learning_rate": [0.001, 0.01,], "max_depth": [5,8,10,12],	learning_rate=0.01, max_depth=5, n_estimators=500	learning_rate=0.01, max_depth=8, n_estimators=500
XGBoost	"n_estimators": [200,300,400,500], "learning_rate": [0.001, 0.01,], "max_depth": [5,8,10,12]	learning_rate=0.01, max_depth=8, n_estimators=500	learning_rate=0.01, max_depth=10, n_estimators=500

4.4 Hybrid models architecture

The predictive framework was designed around stacked generalization (stacking), where multiple base regressors (level-0 learners) feed their predictions into a meta-model (level-1 learner), which learns to optimally combine them.

Four hybrid models (denoted as HM1 to HM4) were constructed with different combinations of base learners, as described in Table 3.

Table 3: Hybrid models architecture

Model Name	Base Regressors	Meta Regressor
HM1	Gradient Boosting, XGBoost	Multivariable Regression
HM2	Extra Trees, LightGBM	
HM3	Gradient Boosting, XGBoost, Extra Trees	
HM4	LightGBM, Gradient Boosting, Extra Trees	

Each hybrid model was trained on default parameters and implemented using the StackingRegressor class, where the base models performed individual learning from the input features, and their outputs were aggregated and passed to a linear regressor.

4.5 Model Evaluation Metrics

All models, including the Voting Regressor, were rigorously evaluated on the unseen test dataset using a comprehensive suite of regression performance metrics:

- R^2 (Coefficient of Determination): Indicates the proportion of variance in the target variable explained by the model.
- MSE (Mean Squared Error): Captures the average of the squared differences between predicted and actual values.
- MAE (Mean Absolute Error): Quantifies the mean absolute difference between predicted and observed values.

5. Results

5.1.1 Performance Evaluation of Ensemble Models for CH_4 Excess Sorption Prediction

Table 4: Ensemble Models Result of CH4 sorption

		R2	*MSE*	*MAE*
ExtraTrees	*Train*	0.926073	0.119493	0.173944
	Test	0.917823	0.123003	0.190255
LightGBM	*Train*	0.97791	0.035705	0.121331
	Test	0.969822	0.045171	0.133412
CatBoost	*Train*	0.963651	0.058754	0.165644
	Test	0.951709	0.072282	0.176494
GradientBoosting	*Train*	0.99502	0.00805	0.059318
	Test	0.979946	0.030017	0.098844
XGBoost	*Train*	0.997261	0.004427	0.04199
	Test	0.975931	0.036026	0.089712

A comparative evaluation of ensemble machine learning models was conducted to predict sorption capacity of CH4 in unconventional reservoirs, with a focus on capturing the nonlinear relationships among petrophysical and geochemical parameters. Table 4 summarizes the performance of five models—Extra Trees, LightGBM, CatBoost, Gradient Boosting, and XGBoost using coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE) on both training and test datasets. LightGBM exhibited the best generalization, achieving an R^2 of 0.9698, an MSE of 0.0452, and an MAE of 0.1334 on the test set, closely followed by Gradient Boosting (R^2 = 0.9799) and XGBoost (R^2 = 0.9759). While Gradient Boosting and XGBoost demonstrated slightly higher training R^2 values (0.9950 and 0.9973, respectively), the modest increase in test error indicates mild overfitting. In contrast, Extra Trees and CatBoost, though accurate, showed comparatively higher prediction errors. The actual versus predicted sorption plot in Figure 7 on the test data further confirms model reliability; models with higher R^2 values displayed minimal deviation from the identity line, indicating strong predictive agreement. These findings underscore the applicability of advanced

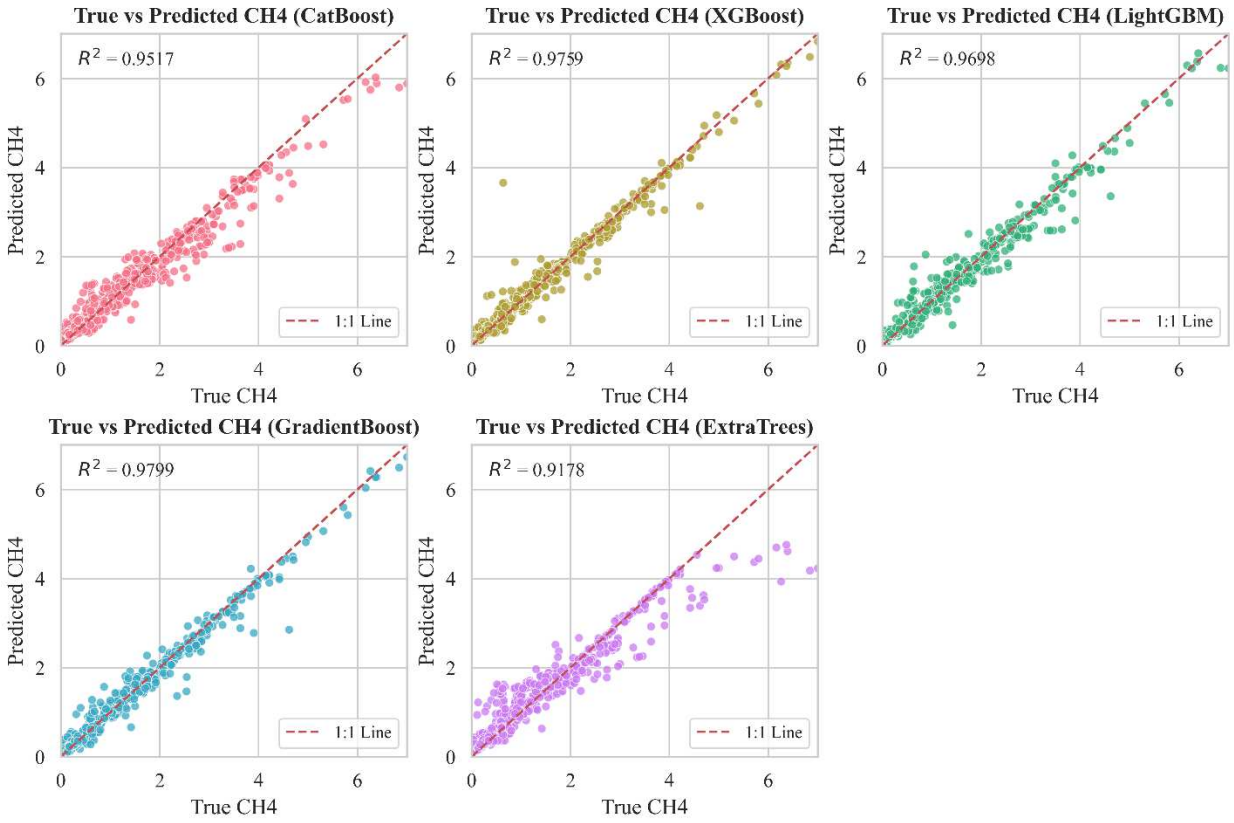


Figure 7: Scatter plot of actual vs predicted CH4 excess sorption

ensemble methods particularly LightGBM in modeling gas adsorption behavior in complex reservoir systems where conventional approaches are often inadequate.

5.1.2 Performance Evaluation of Hybrid Models for CH4 Excess Sorption Prediction

Four high-performing hybrid machine learning models (HM1–HM4) were evaluated for their ability to predict CH_4 sorption capacity in unconventional reservoirs. As shown in Table 5, all models achieved exceptionally high predictive

Table 5: Hybrid Models Result of CH4 sorption

Models		*R2*	*MSE*	*MAE*
HM1	*Train*	0.998135	0.03665	0.003014
	Test	0.983031	0.082436	0.025399
HM2	*Train*	0.999987	0.00024	2.02E-05
	Test	0.990028	0.065192	0.014927
HM3	*Train*	0.999987	0.00024	2.02E-05
	Test	0.990494	0.062706	0.014229
HM4	*Train*	0.999987	0.00024	2.02E-05
	Test	0.990282	0.062691	0.014545

performance, with test R^2 values exceeding 0.98. HM1, while demonstrating a strong fit ($R^2 = 0.9830$, MAE = 0.0254), showed slightly higher test error compared to the other hybrids, suggesting minor overfitting relative to its training performance ($R^2 = 0.9981$). In contrast, HM2, HM3, and HM4 achieved near-perfect training accuracy ($R^2 = 0.999987$, MAE = 2.02×10^{-5}), while maintaining excellent generalization on the test set with R^2 values of 0.9900–0.9905 and consistently low MAEs ranging from 0.0142 to 0.0149. The minimal variance in performance among HM2–HM4 indicates model robustness and reproducibility. The actual versus predicted CH_4 sorption plots for the hybrid modelss on test data in Figure 8 reveal tight clustering around the 1:1 line, confirming high predictive fidelity and minimal residual error. These results highlight the effectiveness of the proposed hybrid architectures in capturing the nonlinear sorption dynamics of CH_4 in low-permeability formations, offering a reliable alternative to laboratory measurements and contributing to improved gas-in-place estimations in shale gas systems.

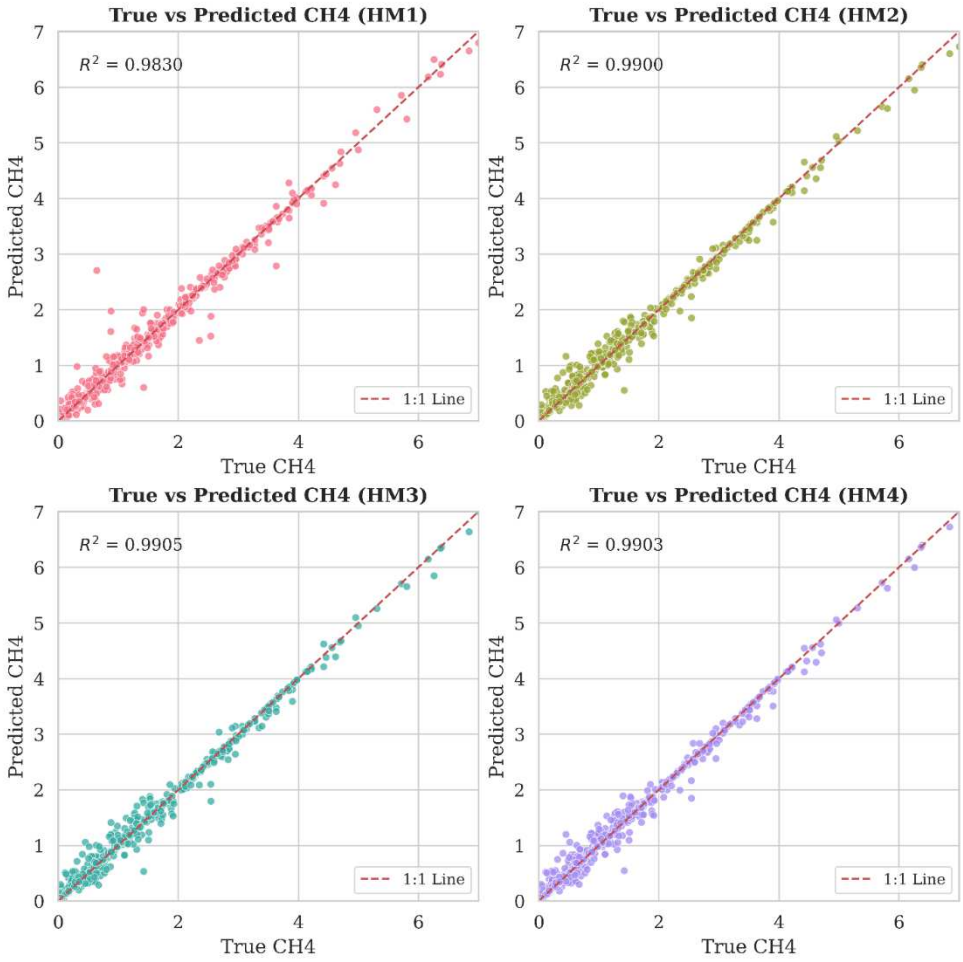


Figure 8: Scatter plot of actual vs predicted CH4 excess sorption (Hybrid models)

5.2.3 Performance Evaluation of Ensemble Models for CO2 Excess Sorption Prediction

Table 6: Ensemble models Result of CO2 sorption

		R2	*MSE*	*MAE*
ExtraTrees	*Train*	0.999841	0.000866	0.006593
	Test	0.994117	0.021932	0.037798
LightGBM	*Train*	0.966673	0.181909	0.098273
	Test	0.973129	0.100168	0.079251
CatBoost	*Train*	0.996436	0.019452	0.040744
	Test	0.994313	0.021199	0.044746
GradientBoosting	*Train*	0.996604	0.018536	0.036569
	Test	0.990239	0.036387	0.046525
XGBoost	*Train*	0.998695	0.007124	0.02251
	Test	0.991856	0.030358	0.042997

Ensemble machine learning algorithms were employed to model CO_2 sorption capacity in unconventional reservoirs, where adsorption mechanisms are governed by intricate geochemical and petrophysical interactions. Table 6 presents the comparative performance of five models Extra Trees, LightGBM, CatBoost, Gradient Boosting, and XGBoost evaluated on both training and test datasets using R^2, MSE, and MAE. Extra Trees achieved the highest test R^2 (0.9941), accompanied by the lowest MAE (0.0378), indicating superior generalization despite near-perfect training performance (R^2 = 0.9998). CatBoost and Gradient Boosting also demonstrated strong predictive capabilities with test R^2 values of 0.9943 and 0.9902, respectively, and low test errors. XGBoost exhibited slightly higher test MSE but maintained a high R^2 of 0.9919 and an MAE of 0.0430, suggesting reliable model stability. In contrast, LightGBM, though still effective (test R^2 = 0.9731), underperformed relative to the others, particularly on training data (R^2 = 0.9667), indicating a less optimal fit. Actual versus predicted plots in Figure 9 for test data confirm that the best-performing models particularly Extra Trees and CatBoost exhibit tight adherence to the identity line, evidencing high predictive accuracy. These findings affirm the applicability of ensemble models, especially tree-based approaches, for accurately capturing CO_2 sorption behavior in complex low-permeability systems, with potential to enhance reservoir simulation and CO_2 storage assessment workflows.

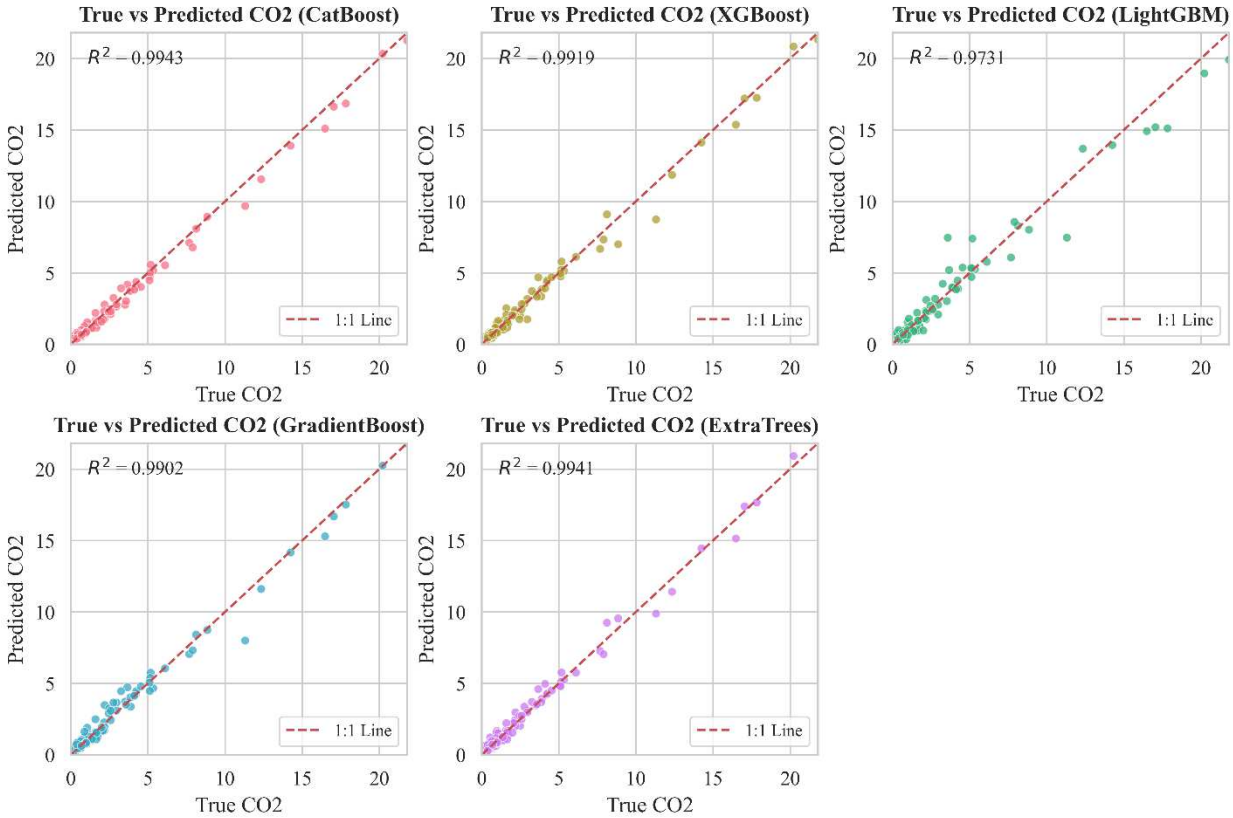


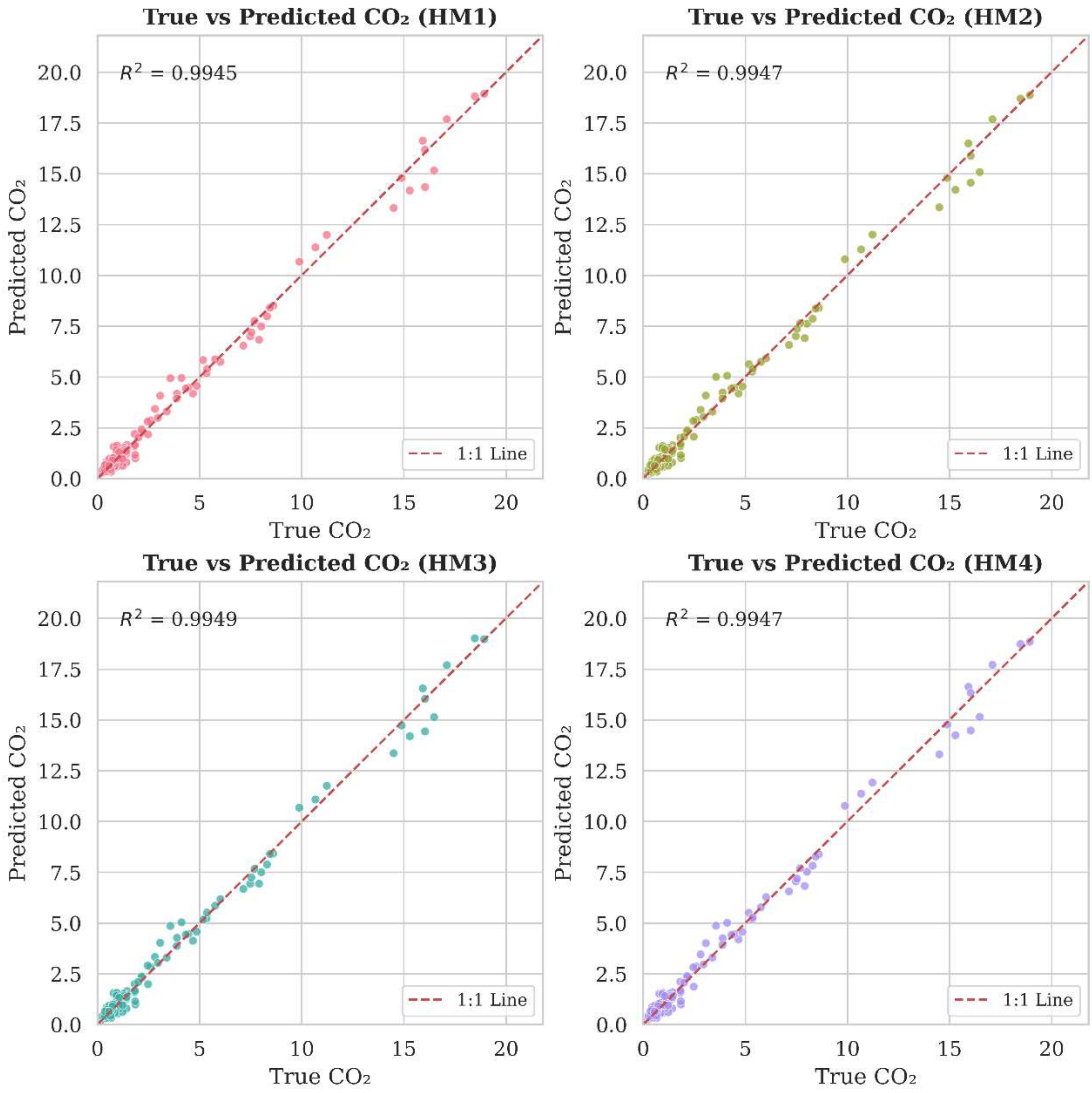
Figure 9: Scatter plot of actual vs predicted CO2 excess sorption (Ensemble models)

5.2.2 Performance Evaluation of Ensemble Models for CO2 Excess Sorption Prediction

As detailed in Table 7, all hybrid models achieved near-perfect performance on the training set ($R^2 = 0.999998$, MSE = 1.68×10^{-4}, MAE = 1.18×10^{-5}), indicating excellent model convergence. On the test data, the models maintained high generalization capability, with R^2 values ranging from 0.9945 to 0.9949 and MAE values between 0.0284 and 0.0308. Among them, HM3 slightly outperformed the others in terms of both R^2 (0.9949) and MAE (0.0284), although all models exhibited minimal variance, highlighting their robustness and reproducibility. The actual versus predicted CO_2 sorption plots in Figure 10 for the test data confirm that predictions closely align with measured values, as evidenced by data points tightly clustering around the identity line. These results demonstrate the hybrid models' efficacy in capturing the intricacies of CO_2 adsorption processes in low-permeability formations and affirm their utility as reliable predictive tools in reservoir evaluation and CO_2 sequestration assessments.

Table 7: Hybrid Models Result of CO2 sorption

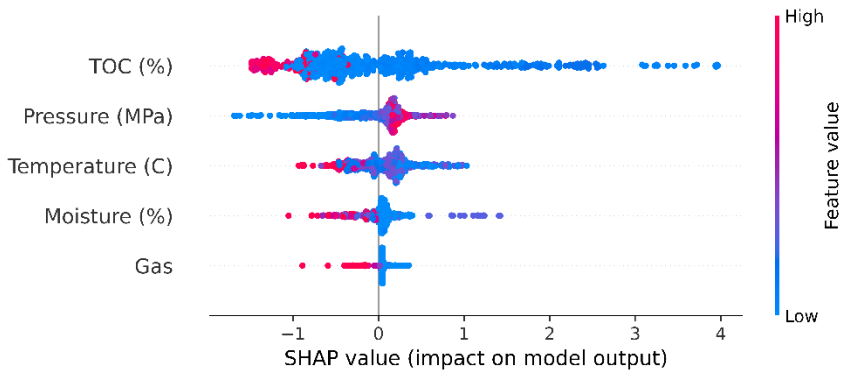
		R2	*MSE*	*MAE*
HM1	*Train*	0.999998	0.000168	1.18E-05
	Test	0.994491	0.047972	0.030787
HM2	*Train*	0.999998	0.000168	1.18E-05
	Test	0.994741	0.046524	0.02939
HM3	*Train*	0.999998	0.000168	1.18E-05
	Test	0.994921	0.046309	0.028385
HM4	*Train*	0.999998	0.000168	1.18E-05
	Test	0.994675	0.047716	0.029759



361 Figure 10: Scatter plot of actual vs predicted CO2 excess sorption (Hybrid models)

362

363 **5.3 Input features ranking using SHAP**



364 Figure 11: Feature importance for CH4 sorption.

For CH_4 sorption, SHAP (SHapley Additive exPlanations) values were utilized to determine the relative importance of input features. Among these, Total Organic Carbon (TOC) was identified as the most influential parameter, followed by pressure, temperature, and moisture content (Figure 11). In the case of CO_2 sorption (Figure 12), SHAP analysis revealed that gas concentration (%) was the most dominant factor, with temperature and pressure also playing significant roles. Notably, the model accurately captured the positive effect of elevated TOC levels on CO_2 sorption, consistent with the findings reported by Klewiah et al. (2019).

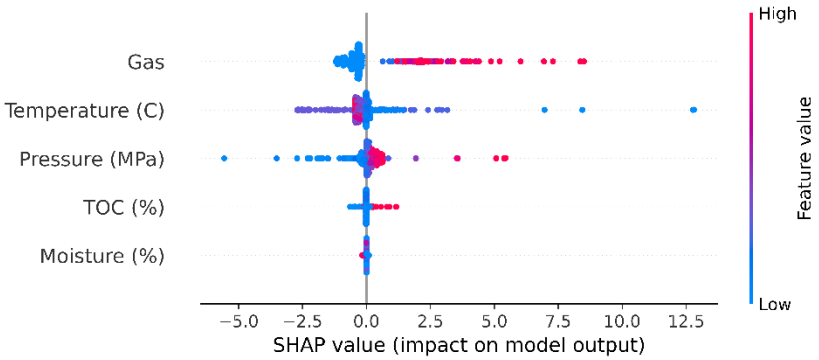


Figure 12: Feature importance of CO2 sorption

5.4 Partial Dependency analysis of CO2 Sorption

Figure 13 illustrates how different reservoir parameters influence CO_2 sorption in unconventional reservoirs based on partial dependence plots. Gas content emerges as the most significant driver, with CO_2 uptake rising almost linearly as gas content increases, indicating its dominant role in sorption capacity. Total Organic Content (TOC) shows a strong effect initially, rapidly boosting CO_2 sorption at low percentages, but its influence levels off at higher concentrations, suggesting diminishing returns. Moisture content exhibits a slight negative impact; as moisture increases, CO_2 sorption decreases marginally, likely due to water occupying available sorption sites. The effect of temperature is pronounced higher temperatures lead to a marked reduction in sorption, which is consistent with the exothermic nature of CO_2 sorption processes. Pressure, meanwhile, enhances CO_2 sorption significantly at lower values but this effect plateaus as pressure increases. Overall, these results highlight that gas content, pressure, and temperature are the primary factors controlling CO_2 storage capacity in unconventional reservoirs, whereas TOC and moisture have more modest effects which are similar to the results from section 5.3.

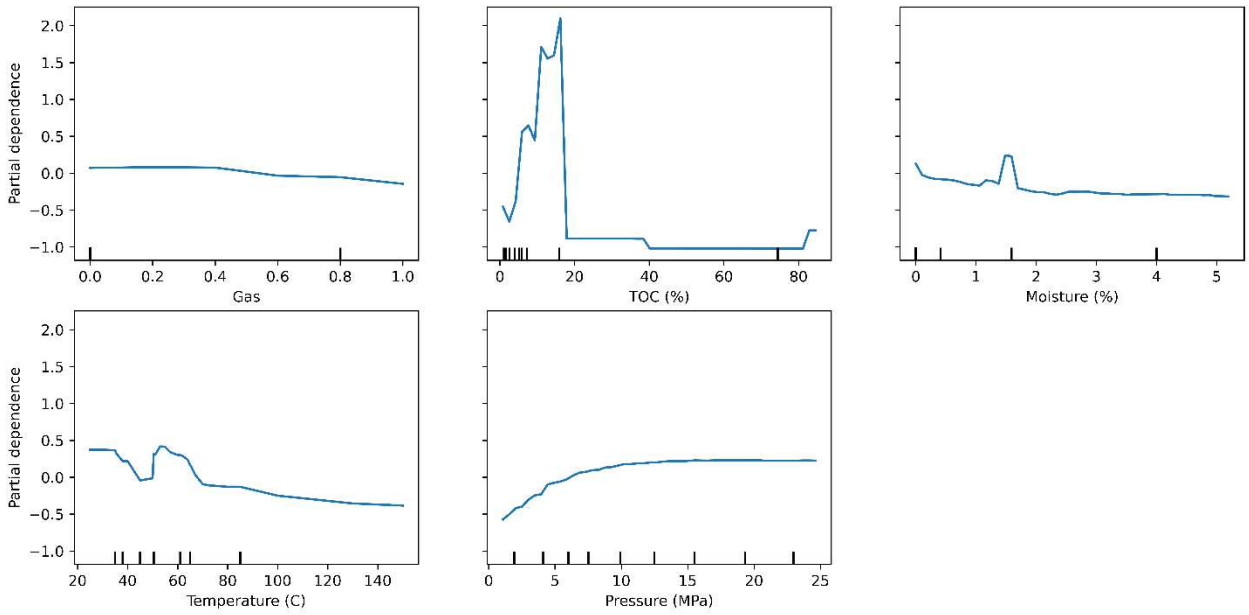


Figure 13: Partial Dependency plot of CO2 Sorption

5.5 Partial Dependency analysis of CH4 Sorption

The partial dependence plots in Figure 14 show how each variable affects methane (CH_4) sorption in unconventional reservoirs. Sorption increases sharply with total organic carbon (TOC) at low values but plateaus at higher TOC, indicating limited benefit beyond a threshold. Higher moisture reduces CH_4 sorption, likely due to competition for

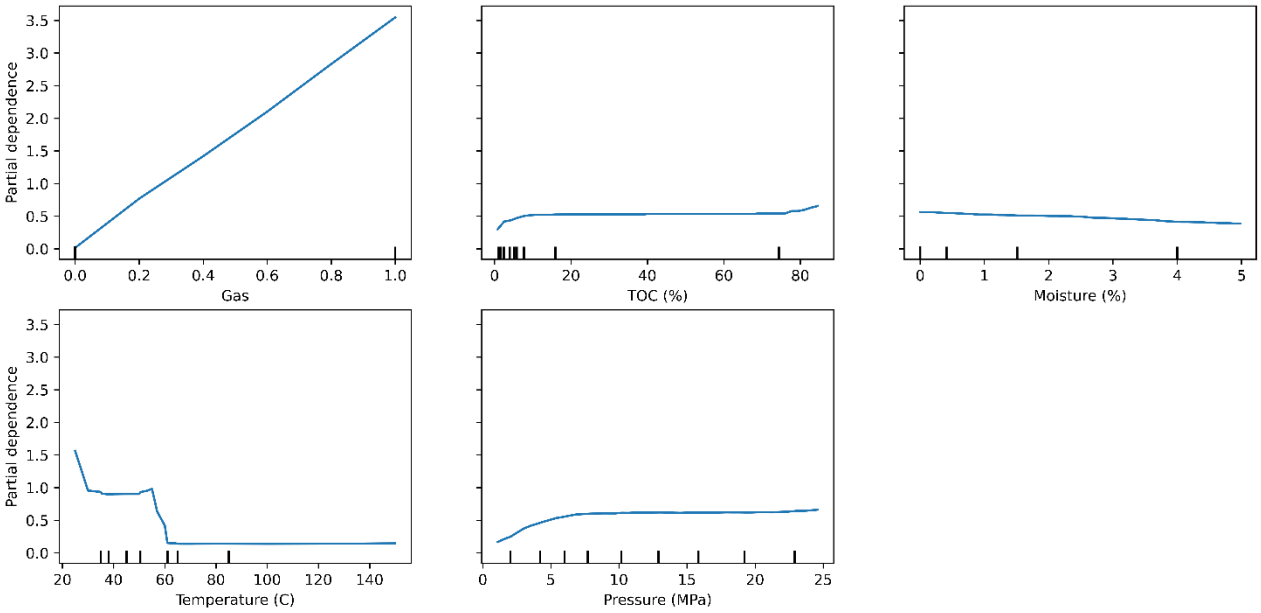


Figure 14: Partial Dependency plot of CH4 Sorption

adsorption sites. Temperature has a mild negative effect higher temperatures reduce sorption. Pressure increases sorption quickly at low values, then levels off, reflecting site saturation. Gas content shows a slight decrease in sorption with higher values, possibly due to saturation effects. Overall, TOC and pressure are the dominant controls, with moisture and temperature as secondary factors whuch are consistent with domain knowledge and complement findings in section 5.3.

5.6 Sorption profile in isotherm for CO_2 and CH_4

This Figure 15 presents a comparison between experimental measurements and model predictions of CO_2 and CH_4 sorption isotherms in coal and shale under different pressures, temperatures, and organic content (TOC). Panels (a)–(d) show single-gas adsorption, where both CO_2 and CH_4 uptake increase with pressure and plateau. The plots demonstrate good model agreement with experiments under various conditions. Panel (e) compares binary sorption of CO_2 and CH_4 in shale, illustrating that CO_2 consistently adsorbs more strongly than CH_4, especially at higher pressures. Overall, the figure highlights the reliability of the predictive model and underscores that pressure, organic content, and gas type are key factors influencing sorption capacity in unconventional reservoirs. Conditions of Figure 15:

- **Panel (a):** Coal, CO_2 sorption at high TOC (79.5), 55°C, pressures up to ~28MPa.
- **Panel (b):** Shale, CO_2 sorption at low TOC (1.8), 35°C, pressures up to 10MPa.
- **Panel (c):** Shale, CH_4 sorption at very low TOC (0.09), 38°C, pressures up to 14MPa.
- **Panel (d):** Shale, CH_4 sorption at moderate TOC (1.66), 50°C, pressures up to 18MPa.
- **Panel (e):** Shale, binary CO_2/CH_4 sorption at TOC (1.8), 35°C, pressures up to 10MPa.

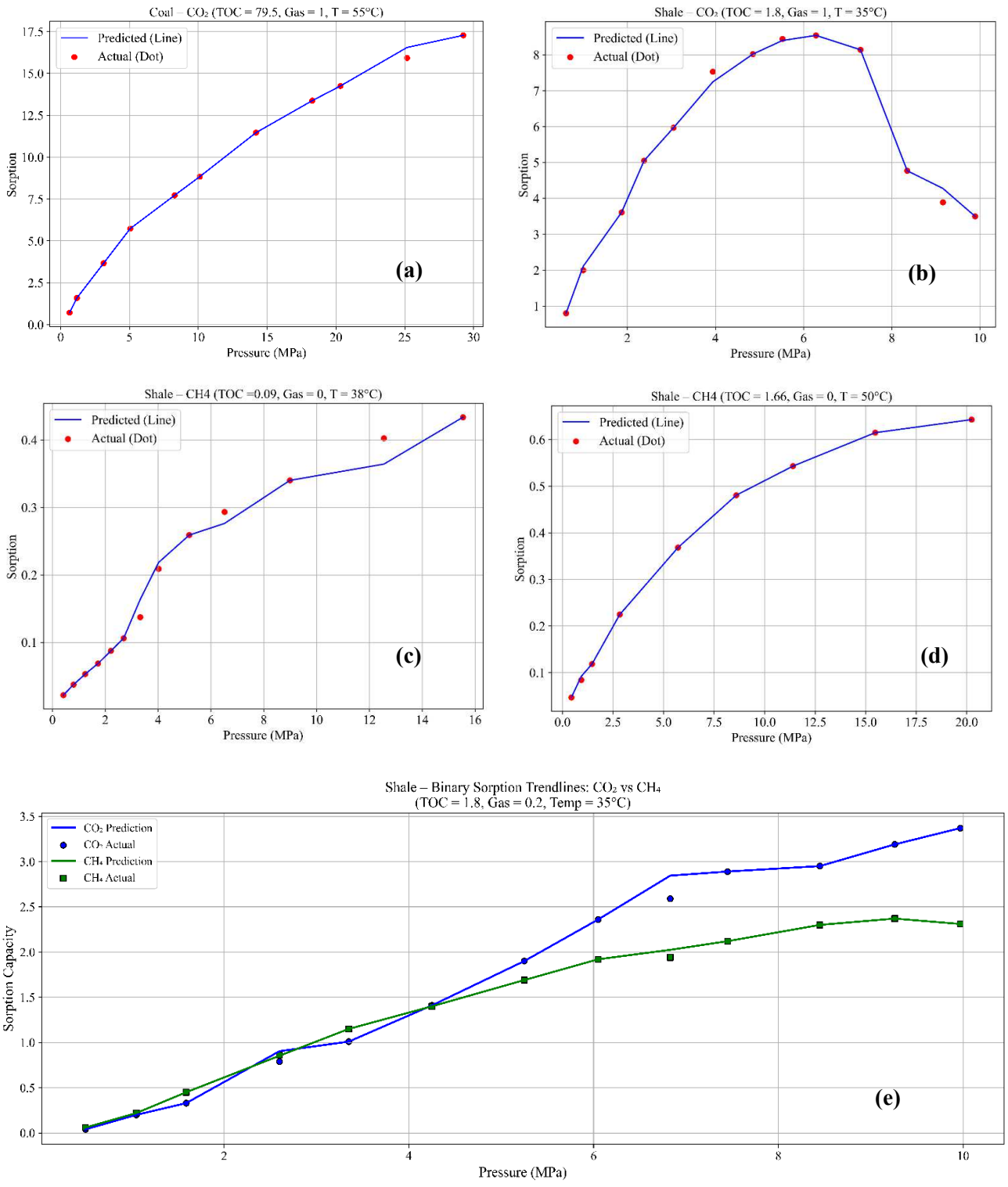
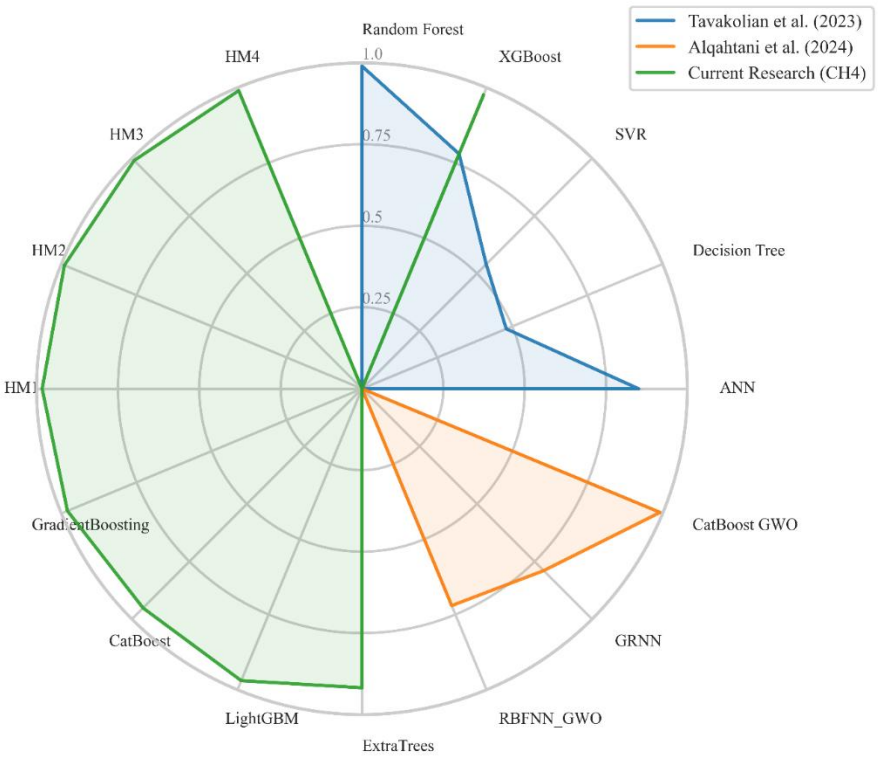


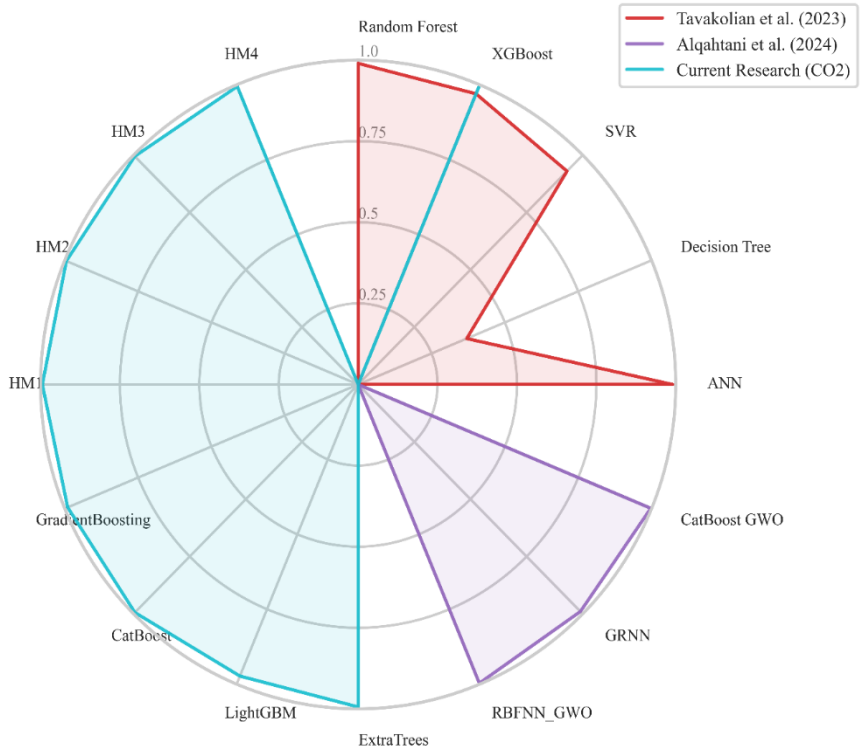
Figure 15: Experimental and Predicted Sorption Isotherms for CO_2 and CH_4 in Coal and Shale Reservoirs

5.7 Comparison with previous study

To contextualize the performance of the proposed models, a comparative analysis was conducted against previously published results by Tavakolian et al. (2023) and Alqahtani et al. (2024) both of which investigated CH_4 and CO_2 sorption capacity prediction in tight reservoirs using machine learning techniques. For CH_4 sorption, Tavakolian et al. reported the highest R^2 value of 0.99 using Random Forest, with other models such as XGBoost (0.78), SVR (0.54), Decision Tree (0.48), and ANN (0.85) showing markedly lower predictive capabilities. Alqahtani et al. (2024) achieved an R^2 of 0.9918 with a CatBoost-GWO hybrid, while GRNN and RBFNN-GWO produced lower scores (0.7896 and 0.7204, respectively). In comparison, our models demonstrated consistently strong performance, with traditional ensemble methods such as LightGBM (R^2 = 0.9698), CatBoost (0.9517), Gradient Boosting (0.9799), and XGBoost (0.9759) outperforming the baseline results of prior works. Notably, our hybrid models (HM1–HM4) exhibited further improvements, achieving R^2 values ranging from 0.9830 to 0.9905, thereby matching or exceeding the best models reported in the literature. A similar trend was observed for CO_2 sorption modeling. While Tavakolian et al. reported R^2 values of 0.99 using both Random Forest and ANN, and Alqahtani et al. (2024) achieved 0.9974 with CatBoost-GWO, our top-performing models including Extra Trees (0.9942), CatBoost (0.9943), and the hybrid models (HM1–HM4: R^2 = 0.9945–0.9949) demonstrated comparable accuracy with marginally reduced variance across test scenarios. The accompanying radar plot (Figure 16 and Figure 17) further substantiates these findings, illustrating the superior and consistent performance of the proposed hybrid models across both gases, with minimal divergence in R^2, MSE, and MAE metrics. Unlike the more scattered performance patterns of individual models reported in earlier studies, our radar visualization highlights the hybrid models' robustness and generalization capability across a wide range of sorption conditions. Collectively, these results affirm that the hybrid approaches developed in this study offer a significant advancement over existing machine learning frameworks for accurate and stable sorption capacity prediction in unconventional reservoirs.



456 Figure 16: Radar plot of CH4 Scores



457 Figure 17: Radar plot of CO2 Scores

5.8 Limitations and Recommendations for future work

A notable limitation of this study concerns the handling of moisture content. In several referenced experimental datasets, rock samples were deliberately dried prior to sorption testing. This pre-treatment deviates from actual reservoir conditions, where in-situ moisture plays a critical role in influencing sorption dynamics. Consequently, the moisture-related findings derived from such datasets may not fully capture the true behavior of reservoir rocks under natural conditions, potentially impacting the reliability and applicability of these insights.

While the study demonstrates the strong predictive capabilities of machine learning models for estimating sorption capacity, it also acknowledges several areas requiring further advancement. A primary limitation lies in the study's predominantly data-driven methodology, which emphasizes statistical correlations between input variables and output responses. Although this approach can achieve high predictive accuracy, it may fail to capture the fundamental physical principles governing sorption phenomena. Bridging this gap calls for the integration of hybrid modeling techniques, such as Physics-Informed Neural Networks (PINNs), which embed known physical laws into the learning process to enhance model interpretability and scientific relevance.

Additionally, the generalizability of machine learning models remains a challenge, particularly when applied to conditions that extend beyond the scope of the training data. This limitation underscores the necessity of improving the extrapolation capabilities of predictive models. Future research could address this by investigating advanced strategies such as transfer learning, domain adaptation, and synthetic data augmentation. These approaches offer promising avenues to enhance model robustness and adaptability in complex, unfamiliar, or extreme reservoir environments.

6. Conclusion

This study developed and validated a comprehensive machine learning framework for predicting CH_4 and CO_2 sorption capacities in tight reservoirs under diverse thermodynamic and geological conditions. All models demonstrated high predictive performance; however, the hybrid models proposed in this work consistently outperformed conventional approaches. Key conclusions are as follows:

- **Model Accuracy:** Hybrid Model 3 (HM3) achieved the highest predictive accuracy for both CH_4 ($R^2 = 0.9905$) and CO_2 ($R^2 = 0.9949$), surpassing previously reported results in the literature.

- **Feature Importance:** SHAP and partial dependence analyses identified total organic carbon (TOC) and pressure as dominant controls on CH_4 sorption. For CO_2, GAS(%) representing CO_2 concentration was the most influential, followed by temperature and pressure.
- **Reservoir Behavior:** The models accurately captured key adsorption dynamics, including the decline in CO_2 sorption at elevated pressures and the saturation effects of TOC and pressure, aligning with known reservoir phenomena.
- **Model Robustness:** All models maintained strong performance across varying input conditions, with minimal overfitting, demonstrating excellent generalization.
- **Novel Contribution:** This work introduces the application of hybrid ensemble models (HM1–HM4), along with ExtraTrees, CatBoost, and LightGBM, to sorption prediction—expanding the modeling toolkit for unconventional reservoirs.
- **Future Work:** Incorporating additional parameters such as porosity, specific surface area, and salinity, and coupling these models with reservoir simulators could enhance predictive fidelity and support operational decis
- ion-making.

This study confirms the potential of ML models to deliver fast, reliable predictions of sorption capacity, thereby reducing dependency on expensive laboratory measurements and supporting more efficient reservoir management strategies.

Data availability statement

Data will be available upon request.

Credit authorship contribution statement

Abu Bakker Siddique: Writing – original draft, Software. Minhaz Chowdhury: Writing – original draft, Software. Mahamudul Hashan: Writing – review & editing, Writing – original draft, Supervision, Conceptualization. Labiba Nusrat Jahan: Writing – original draft, Software. Md. Iftiar Uddin: Writing – original draft, Software.

Declaration of Competing Interest

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

Acknowledgment

The authors acknowledge the Shahjalal University of Science and Technology (SUST) Research Center for providing funding to carry out this study.

References

Beutekamp, S., & Harting, P. (2002). Experimental Determination and Analysis of High Pressure Adsorption Data of Pure Gases and Gas Mixtures. *Adsorption*, *8*(4), 255–269. https://doi.org/10.1023/a:1021548112040

King, C. J. (2013). *Separation processes*. Courier Corporation.

Goetz, V., Pupier, O., & Guillot, A. (2006). Carbon dioxide-methane mixture adsorption on activated carbon. *Adsorption*, *12*(1), 55–63. https://doi.org/10.1007/s10450-006-0138-z

Dong, A., Xie, J., Wang, W., Yu, L., Liu, Q., & Yin, Y. (2010). A novel method for amino starch preparation and its adsorption for Cu(II) and Cr(VI). *Journal of Hazardous Materials*, *181*(1–3), 448–454. https://doi.org/10.1016/j.jhazmat.2010.05.031

Rebouh, S., Bouhedda, M., & Hanini, S. (2015). Neuro-fuzzy modeling of Cu(II) and Cr(VI) adsorption from aqueous solution by wheat straw. *Desalination and Water Treatment*, *57*(14), 6515–6530. https://doi.org/10.1080/19443994.2015.1009171

Chalmers, G. R., & Bustin, R. M. (2008). Lower Cretaceous gas shales in northeastern British Columbia, Part I: geological controls on methane sorption capacity. *Bulletin of Canadian Petroleum Geology*, *56*(1), 1–21. https://doi.org/10.2113/gscpgbull.56.1.1

529 Ross, D. J. K., & Bustin, R. M. (2007). Characterizing the shale gas resource potential of Devonian–Mississippian
530 strata in the Western Canada sedimentary basin: Application of an integrated formation evaluation. *AAPG Bulletin*,
531 *92*(1), 87–125. https://doi.org/10.1306/09040707048

532 Yang, T., Nie, B., Yang, D., Zhang, R., & Zhao, C. (2011). Experimental research on displacing coal bed methane
533 with supercritical CO2. *Safety Science*, *50*(4), 899–902. https://doi.org/10.1016/j.ssci.2011.08.011

534 Yang, Y., & Liu, S. (2020). Review of shale gas sorption and its models. *Energy & Fuels*, *34*(12), 15502–15524.
535 https://doi.org/10.1021/acs.energyfuels.0c02906

536 Martini, A., Walter, L., Budai, J., Ku, T., Kaiser, C., & Schoell, M. (1998). Genetic and temporal relations between
537 formation waters and biogenic methane: Upper Devonian Antrim Shale, Michigan Basin, USA. *Geochimica Et*
538 *Cosmochimica Acta*, *62*(10), 1699–1720. https://doi.org/10.1016/s0016-7037(98)00090-8

539 Hill, R. J., Zhang, E., Katz, B. J., & Tang, Y. (2007). Modeling of gas generation from the Barnett Shale, Fort Worth
540 Basin, Texas. *AAPG Bulletin*, *91*(4), 501–521. https://doi.org/10.1306/1206060606

541 Bernard, S., Horsfield, B., Schulz, H., Schreiber, A., Wirth, R., Vu, T. T. A., Perssen, F., Könitzer, S., Volk, H.,
542 Sherwood, N., & Fuentes, D. (2010). Multi-scale detection of organic and inorganic signatures provides insights into
543 gas shale properties and evolution. *Geochemistry*, *70*, 119–133. https://doi.org/10.1016/j.chemer.2010.05.005

544 Glorioso, J. C., & Rattia, A. (2012). Unconventional reservoirs: basic petrophysical concepts for shale gas. *All Days*.
545 https://doi.org/10.2118/153004-ms

546 S.R. Reeves An overview of CO2-ECBM and sequestration in coal seams. 2009.

547 Mukherjee, M., & Misra, S. (2018). A review of experimental research on Enhanced Coal Bed Methane (ECBM)
548 recovery via CO2 sequestration. *Earth-Science Reviews*, *179*, 392–410.
549 https://doi.org/10.1016/j.earscirev.2018.02.018

550 Busch, A., & Gensterblum, Y. (2011). CBM and CO2-ECBM related sorption processes in coal: A review.
551 *International Journal of Coal Geology*, *87*(2), 49–71. https://doi.org/10.1016/j.coal.2011.04.011

552 Kang, S. M., Fathi, E., Ambrose, R. J., Akkutlu, I. Y., & Sigal, R. F. (2010). Carbon dioxide storage capacity of
553 organic-rich shales. *SPE Annual Technical Conference and Exhibition*. https://doi.org/10.2118/134583-ms

554 Zhang, X. G., Ranjith, P. G., Perera, M. S. A., Ranathunga, A. S., & Haque, A. (2016). Gas transportation and
555 enhanced coalbed methane recovery processes in deep coal seams: a review. *Energy & Fuels*, *30*(11), 8832–8849.
556 https://doi.org/10.1021/acs.energyfuels.6b01720

557 Liu, S. (2020). What is bioprocess engineering? In *Bioprocess Engineering* (pp. 1–15). https://doi.org/10.1016/b978-
558 0-12-821012-3.00001-4

559 Liu, F., Ellett, K., Xiao, Y., & Rupp, J. A. (2013). Assessing the feasibility of CO2 storage in the New Albany Shale
560 (Devonian–Mississippian) with potential enhanced gas recovery using reservoir simulation. *International Journal of*
561 *Greenhouse Gas Control*, *17*, 111–126. https://doi.org/10.1016/j.ijggc.2013.04.018

562 R. Aguilera Unconventional gas and tight oil exploitation. SPE Richardson. 2018.

563 Krooss, B., Van Bergen, F., Gensterblum, Y., Siemons, N., Pagnier, H., & David, P. (2002). High-pressure methane
564 and carbon dioxide adsorption on dry and moisture-equilibrated Pennsylvanian coals. *International Journal of Coal*
565 *Geology*, *51*(2), 69–92. https://doi.org/10.1016/s0166-5162(02)00078-2

566 Fitzgerald, J., Pan, Z., Sudibandriyo, M., Robinson, R., Jr, Gasem, K., & Reeves, S. (2005). Adsorption of methane,
567 nitrogen, carbon dioxide and their mixtures on wet Tiffany coal. *Fuel*, *84*(18), 2351–2363.
568 https://doi.org/10.1016/j.fuel.2005.05.002

569 Li, D., Liu, Q., Weniger, P., Gensterblum, Y., Busch, A., & Krooss, B. M. (2009). High-pressure sorption isotherms
570 and sorption kinetics of CH4 and CO2 on coals. *Fuel*, *89*(3), 569–580. https://doi.org/10.1016/j.fuel.2009.06.008

571 Zhang, S., Tang, S., Li, Z., Pan, Z., & Liu, B. (2020). Competitive sorption and diffusion of methane and carbon
572 dioxide mixture in Carboniferous-Permian anthracite of south Qinshui Basin, China. *Arabian Journal of Geosciences*,
573 *13*(24). https://doi.org/10.1007/s12517-020-06303-9

574 Zhang, Q., Zhu, H., & Kang, R. (2023). Influence of uniaxial strain loading on the adsorption-diffusion properties of
575 binary components of CH4/CO2 in micropores of bituminous coal by macromolecular simulation. *Powder*
576 *Technology*, *427*, 118715. https://doi.org/10.1016/j.powtec.2023.118715

577 Crosdale, P. J., Beamish, B., & Valix, M. (1998). Coalbed methane sorption related to coal composition. *International*
578 *Journal of Coal Geology*, *35*(1–4), 147–158. https://doi.org/10.1016/s0166-5162(97)00015-3

Lu, X., Li, F., & Watson, A. T. (1995). Adsorption studies of natural gas storage in Devonian shales. *SPE Formation Evaluation*, *10*(02), 109–113. https://doi.org/10.2118/26632-pa

Chalmers, G. R., & Bustin, R. M. (2006). The organic matter distribution and methane capacity of the Lower Cretaceous strata of Northeastern British Columbia, Canada. *International Journal of Coal Geology*, *70*(1–3), 223–239. https://doi.org/10.1016/j.coal.2006.05.001

Chalmers, G. R., & Bustin, R. M. (2008). Lower Cretaceous gas shales in northeastern British Columbia, Part I: geological controls on methane sorption capacity. *Bulletin of Canadian Petroleum Geology*, *56*(1), 1–21. https://doi.org/10.2113/gscpgbull.56.1.1

Ross, D. J. K., & Bustin, R. M. (2007). Characterizing the shale gas resource potential of Devonian–Mississippian strata in the Western Canada sedimentary basin: Application of an integrated formation evaluation. *AAPG Bulletin*, *92*(1), 87–125. https://doi.org/10.1306/09040707048

Ross, D. J., & Bustin, R. M. (2008). The importance of shale composition and pore structure upon gas storage potential of shale gas reservoirs. *Marine and Petroleum Geology*, *26*(6), 916–927. https://doi.org/10.1016/j.marpetgeo.2008.06.004

Weniger, P., Kalkreuth, W., Busch, A., & Krooss, B. M. (2010). High-pressure methane and carbon dioxide sorption on coal and shale samples from the Paraná Basin, Brazil. *International Journal of Coal Geology*, *84*(3–4), 190–205. https://doi.org/10.1016/j.coal.2010.08.003

Zhang, T., Ellis, G. S., Ruppel, S. C., Milliken, K., & Yang, R. (2012). Effect of organic-matter type and thermal maturity on methane adsorption in shale-gas systems. *Organic Geochemistry*, *47*, 120–131. https://doi.org/10.1016/j.orggeochem.2012.03.012

Gasparik, M., Bertier, P., Gensterblum, Y., Ghanizadeh, A., Krooss, B. M., & Littke, R. (2013). Geological controls on the methane storage capacity in organic-rich shales. *International Journal of Coal Geology*, *123*, 34–51. https://doi.org/10.1016/j.coal.2013.06.010

Pan, Z., Connell, L. D., Camilleri, M., & Connelly, L. (2010). Effects of matrix moisture on gas diffusion and flow in coal. *Fuel*, *89*(11), 3207–3217. https://doi.org/10.1016/j.fuel.2010.05.038

Gensterblum, Y., Merkel, A., Busch, A., & Krooss, B. M. (2013). High-pressure CH4 and CO2 sorption isotherms as a function of coal maturity and the influence of moisture. *International Journal of Coal Geology*, *118*, 45–57. https://doi.org/10.1016/j.coal.2013.07.024

Rother, G., Krukowski, E. G., Wallacher, D., Grimm, N., Bodnar, R. J., & Cole, D. R. (2011). Pore size effects on the sorption of supercritical CO2 in mesoporous CPG-10 silica. *The Journal of Physical Chemistry C*, *116*(1), 917–922. https://doi.org/10.1021/jp209341q

Shabani, M., Moallemi, S. A., Krooss, B. M., Amann-Hildenbrand, A., Zamani-Pozveh, Z., Ghalavand, H., & Littke, R. (2017). Methane sorption and storage characteristics of organic-rich carbonaceous rocks, Lurestan province, southwest Iran. *International Journal of Coal Geology*, *186*, 51–64. https://doi.org/10.1016/j.coal.2017.12.005

Arunachalam, K. (2021). *Adsorption and migration mechanisms of methane and carbon dioxide in gas shales at pore scale*. https://doi.org/10.7488/era/1062

Zhang, X. G., Ranjith, P. G., Perera, M. S. A., Ranathunga, A. S., & Haque, A. (2016b). Gas transportation and enhanced coalbed methane recovery processes in deep coal seams: a review. *Energy & Fuels*, *30*(11), 8832–8849. https://doi.org/10.1021/acs.energyfuels.6b01720

Gasparik, M., Ghanizadeh, A., Bertier, P., Gensterblum, Y., Bouw, S., & Krooss, B. M. (2012). High-Pressure Methane Sorption Isotherms of Black Shales from The Netherlands. *Energy & Fuels*, *26*(8), 4995–5004. https://doi.org/10.1021/ef300405g

Ji, L., Zhang, T., Milliken, K. L., Qu, J., & Zhang, X. (2012). Experimental investigation of main controls to methane adsorption in clay-rich rocks. *Applied Geochemistry*, *27*(12), 2533–2545. https://doi.org/10.1016/j.apgeochem.2012.08.027

Cao, M., Qin, C., Jiang, Y., Xia, P., & Wang, K. (2025). Review of Competitive Adsorption of CO2/CH4 in shale: Implications for CO2 sequestration and Enhancing shale gas Recovery. *ACS Omega*. https://doi.org/10.1021/acsomega.4c08678

Hui, D., Li, L., Zhang, Y., Peng, X., Li, T., Jia, C., & Pan, Y. (2023). Molecular simulation of adsorption behaviors of

methane and carbon dioxide on typical clay minerals. *Frontiers in Energy Research*, *11*. https://doi.org/10.3389/fenrg.2023.1231338

Busch, A., Gensterblum, Y., & Krooss, B. M. (2003). Methane and CO2 sorption and desorption measurements on dry Argonne premium coals: pure components and mixtures. *International Journal of Coal Geology*, *55*(2–4), 205–224. https://doi.org/10.1016/s0166-5162(03)00113-7

Busch, A., Gensterblum, Y., Krooss, B. M., & Littke, R. (2004). Methane and carbon dioxide adsorption–diffusion experiments on coal: upscaling and modeling. *International Journal of Coal Geology*, *60*(2–4), 151–168. https://doi.org/10.1016/j.coal.2004.05.002

Wang, K., Wang, G., Ren, T., & Cheng, Y. (2014). Methane and CO 2 sorption hysteresis on coal: A critical review. *International Journal of Coal Geology*, *132*, 60–80. https://doi.org/10.1016/j.coal.2014.08.004

Zhang, S., Tang, S., Li, Z., Liu, B., & Wang, R. (2022). Effect of pore structure on competitive sorption and diffusion of mixed methane and carbon dioxide in anthracite, South Qinshui Basin, China. *International Journal of Coal Geology*, *253*, 103956. https://doi.org/10.1016/j.coal.2022.103956

Bhattacharya, S. (2021). A primer on machine learning in subsurface geosciences. In *SpringerBriefs in petroleum geoscience & engineering*. https://doi.org/10.1007/978-3-030-71768-1

Okoroafor, E. R., Smith, C. M., Ochie, K. I., Nwosu, C. J., Gudmundsdottir, H., & Aljubran, M. (2022). Machine learning in subsurface geothermal energy: Two decades in review. *Geothermics*, *102*, 102401. https://doi.org/10.1016/j.geothermics.2022.102401

Samnioti, A., & Gaganis, V. (2023). Applications of Machine Learning in Subsurface Reservoir Simulation—A Review—Part II. *Energies*, *16*(18), 6727. https://doi.org/10.3390/en16186727

Liao, Q., Zhou, J., Zheng, Y., Xian, X., Deng, G., Zhang, C., Duan, X., Wu, Z., & Li, S. (2024). Adsorption kinetics of CH4 and CO2 on shale: Implication for CO2 sequestration. *Separation and Purification Technology*, *337*, 126427. https://doi.org/10.1016/j.seppur.2024.126427

Liao, Q., Zhou, J., Xian, X., Yang, K., Zhang, C., Dong, Z., & Yin, H. (2023). Competition adsorption of CO2/CH4 in shale: Implications for CO2 sequestration with enhanced gas recovery. *Fuel*, *339*, 127400. https://doi.org/10.1016/j.fuel.2023.127400

Hussin, F., Rahim, S. a. N. M., Hatta, N. S. M., Aroua, M. K., & Mazari, S. A. (2023). A systematic review of machine learning approaches in carbon capture applications. *Journal of CO2 Utilization*, *71*, 102474. https://doi.org/10.1016/j.jcou.2023.10247

Rostami, A., Anbaz, M. A., Gahrooei, H. R. E., Arabloo, M., & Bahadori, A. (2017). Accurate estimation of CO 2 adsorption on activated carbon with multi-layer feed-forward neural network (MLFNN) algorithm. *Egyptian Journal of Petroleum*, *27*(1), 65–73. https://doi.org/10.1016/j.ejpe.2017.01.003

Barki, H., Khaouane, L., & Hanini, S. (2019). Modelling of adsorption of methane, nitrogen, carbon dioxide, their binary mixtures, and their ternary mixture on activated carbons using artificial neural network. *Kemija U Industriji*, *68*(7–8), 289–302. https://doi.org/10.15255/kui.2019.002

Zhou, L. (2018). Prediction of CO2 adsorption on different activated carbons by hybrid group method of data-handling networks and LSSVM. *Energy Sources Part a Recovery Utilization and Environmental Effects*, *41*(16), 1960–1971. https://doi.org/10.1080/15567036.2018.1548521

Bemani, A., Baghban, A., Mohammadi, A. H., & Andersen, P. Ø. (2020). Estimation of adsorption capacity of CO2, CH4, and their binary mixtures in Quidam shale using LSSVM: Application in CO2 enhanced shale gas recovery and CO2 storage. *Journal of Natural Gas Science and Engineering*, *76*, 103204. https://doi.org/10.1016/j.jngse.2020.103204

Huang, M., Xu, H., Yu, H., Zhang, H., Micheal, M., Yuan, X., & Wu, H. (2022). Fast prediction of methane adsorption in shale nanopores using kinetic theory and machine learning algorithm. *Chemical Engineering Journal*, *446*, 137221. https://doi.org/10.1016/j.cej.2022.137221

Alanazi, A., Ibrahim, A. F., Bawazer, S., Elkatatny, S., & Hoteit, H. (2023). Machine learning framework for estimating CO2 adsorption on coalbed for carbon capture, utilization, and storage applications. *International Journal of Coal Geology*, *275*, 104297. https://doi.org/10.1016/j.coal.2023.104297

Tavakolian, M., Najafi-Silab, R., Chen, N., & Kantzas, A. (2023). Modeling of methane and carbon dioxide sorption capacity in tight reservoirs using Machine learning techniques. *Fuel*, *360*, 130578. https://doi.org/10.1016/j.fuel.2023.130578

Alqahtani, F. M., Youcefi, M. R., Djema, H., Amar, M. N., & Ghasemi, M. (2024). Data-driven framework for predicting the sorption capacity of carbon dioxide and methane in tight reservoirs. *Greenhouse Gases Science and Technology*. https://doi.org/10.1002/ghg.2318

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3-42. doi: 10.1007/s10994-006-6226-1

Ke, G., Meng, Q., & Cai, W. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154. doi: 10.5555/3295222.3295238

Prokhorenkova, L., Gusev, G., & Vorobev, A. (2018). CatBoost: Unbiased boosting with categorical features. *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 6639–6649.doi: 10.5555/3327798.3327893

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. doi: 10.1145/2939672.2939785

Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2), 1-39. doi: 10.1007/s10462-009-9124-7

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Klewiah, I., Berawala, D. S., Walker, H. C. A., Andersen, P. Ø., & Nadeau, P. H. (2019). Review of experimental sorption studies of CO2 and CH4 in shales. *Journal of Natural Gas Science and Engineering*, *73*, 103045. https://doi.org/10.1016/j.jngse.2019.103045

Wolpert, D. H. Stacked generalization. Neural Netw. 5(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1 (1992).

Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Vol. 2 1–758 (Springer, 2009). https://doi.org/10.1007/978-0-387-21606-5

Breiman, L. Stacked regressions. Mach. Learn. 24(1), 49–64. https://doi.org/10.1007/BF00117832 (1996).