DOI: https://doi.org/10.68732/2bb4c53fe078a78b

ISBN: 978-9989-2053-7-8

Modeling of Carbon Dioxide in Unconventional Reservoir using Machine Learning

A. Tripathy
Indian Institute of Technology (ISM) Dhanbad, Dhanbad, Jharkhand, 826004, India
0000-0001-7253-2960

A. Kumar
Indian Institute of Technology (ISM) Dhanbad, Dhanbad, Jharkhand, 826004, India
21je0186@iitism.ac.in
0009-0007-3532-1209

M. Mazumder
Indian Institute of Technology (ISM) Dhanbad, Dhanbad, Jharkhand, 826004, India
0009-0005-6020-3656

Abstract

Carbon Capture, Utilization, and Storage (CCUS) has turned out to be a serious measure to the reduction of worldwide greenhouse gases.CO_2-Enhanced Coalbed Methane (ECBM) recovery has dual advantage as it captures CO_2 and increases methane yields. Conventional experiment techniques to measure adsorption isotherms are time-consuming and expensive. Present investigation gives a strong suggestion to forecast the capacity of CO_2 adsorption using machine learning (ML) algorithm. This study includes the database of 1,101 coal samples representing different coalfields, considering the key physicochemical parameters including moisture content, ash yield, volatile matter, fixed carbon, vitrinite reflectance, and total organic carbon (TOC), pressure and temperature as input. The Random Forest (RF) model was developed using an 80:20 training–testing data split and optimized through 10-fold cross-validation combined with grid-based hyperparameter tuning. The optimal RF configuration consisted of 50 decision trees with a maximum tree depth of 30, where bootstrap aggregation was used so that each tree was trained on a randomly sampled subset of the training data. The small difference between training and testing R^2 values (<0.03), together with low training errors (root mean Squared Error (RMSE) = 0.0268 mmol/g, mean absolute error (MAE) = 0.0143 mmol/g), indicates minimal overfitting and strong generalization capability. Finally predictive performance of the RF model evaluated on four unseen samples data. The model showed great predictive ability with Coefficient of Determination (R^2) of more than 0.97 and RMSE less than 0.073 mmol/g with MAE (0.04).These results confirm that the Random Forest algorithm is a reliable and computationally efficient method for predicting CO_2 adsorption capacity based on the physiochemical properties of coal. This forecasting model can significantly assist rock engineers and geologists in the preliminary evaluation and design of CO_2 sequestration in coal seams.

Keywords

Random Forest, Adsorption capacity, CCUS, ECBM, Coal, ANN.

1 Introduction

CBMThe global transition toward low-carbon energy systems has increased the importance of Carbon Capture, Utilization, and Storage (CCUS) technologies for mitigating greenhouse gas emissions (Ahmed et al., 2020; Jiang et al., 2022). Coal seams are considered favorable reservoirs because of their high capacity for gas adsorption. CO_2-Enhanced Coalbed Methane (ECBM) recovery is an important CCUS application offering a dual advantage: the injected CO_2 selectively adsorbs on coal surfaces, displacing methane that can be recovered as an energy resource, and at the same time allowing for the permanent storage of CO_2.

The CO_2 adsorption behaviour in coal is influenced by several physicochemical properties, including pressure, temperature, moisture content, ash yield, volatile matter, fixed carbon, vitrinite reflectance, and total organic carbon (TOC). The adsorption capacity is often determined by laboratory experiments and characterized by the Langmuir isotherm model (Langmuir, 1918). However, these experimental methods are often expensive and time-consuming, especially when large sample sets must be evaluated. Furthermore, adsorption behaviour in coal involves complex, nonlinear interactions among multiple parameters that are difficult to capture using simple empirical correlations. Recent advancement in machine learning (ML) offers new opportunities for modelling such complex systems. In this study, a Random Forest (RF) model is applied to predict CO_2 adsorption capacity using physicochemical parameters compiled from published experimental datasets, with the objective of developing a reliable and computationally efficient tool for rapid screening of coal reservoirs for CCUS and ECBM applications.

2 Methodology

2.1 Dataset and Input Features

The dataset comprises 1101 CO_2 adsorption measurements from studies of European coalfields (Gensterblum et al., 2010; Weniger et al., 2012), encompassing a broad spectrum of coal ranks and experimental conditions. Eight physicochemical parameters - moisture content, ash yield, volatile matter, fixed carbon, vitrinite reflectance, TOC, pressure, and temperature were used as inputs, with CO_2 adsorption capacity (mmol/g) as the target variable.

2.2 Model Development and Optimization

Before training the model, all input features were standardized using Z-score normalization to eliminate the bias in the scale of the variables. The dataset was randomly divided into training (80%) and testing (20%) groups, in which the training set was only used for model development and hyperparameter optimization, and the test set was only used for the final assessment of model performance on unseen data. In this study, three ML models were developed and compared, i.e., RF, SVR, and ANN. RF is an ensemble method that combines a number of decision trees to avoid overfitting; SVR is a kernel-based procedure to capture nonlinear relationships in high-dimensional feature space; and ANN is a feed-forward network trained by back propagation, which is able to learn complex input-output relationships through inter-connected hidden layers. The optimal hyperparameters for all three models (Table 1) were determined by combining grid search hyperparameter tuning with 10-fold cross-validation, and the predictive performance was evaluated using R2, RMSE, and MAE.

Table 1. Hyperparameter tuning ranges and optimal values for the machine learning models.

Model	**Hyperparameter name**	**Range**	**Optimal hyperparameters**
RF	n-estimators	50, 100, 200	50
	Max depth	10, 20, 30	30
	Min sample split	2, 5, 10	2
	Min sample leaf	1, 2, 4	1
	Max features	Sqrt, log2	log2

ANN	Hidden layer sizes	(50,), (100,), (50, 50), (100, 50)	(50, 50)
	activation	relu, tanh	tanh
	solver	adam, lbfgs	lbfgs
	alpha	0.0001, 0.01	0.01
	learning_rate_init	0.001, 0.01	0.001
SVR	C	(1-1000) log space	464.14
	Gamma	Scale, 10^{-1}, 10^{-2}, 10^{-3}	Scale
	Epsilon	10^{-2}, 10^{-3}, 0.05,10^{-1}	10^{-3}
	Kernal	Rbf, poly	rbf

3 Result

3.1 Model Performance

The four statistical indicators used to assess the performance of the three optimized ML models were R2, MAE, MSE and RMSE. All these metrics with hyperparameter optimization, are shown in Table 2.

Table 2. Evaluation metrics for ML models with hyperparameter optimization.

Model	**R^2 (Train)**	**MSE (Train)**	**RMSE (Train)**	**MAE (Train)**	**R^2 (Test)**	**MSE (Test)**	**RMSE (Test)**	**MAE (Test)**
RF	0.9953	0.0007	0.0268	0.0143	0.9700	0.0054	0.0732	0.0411
ANN	0.9833	0.0026	0.0506	0.0267	0.9633	0.0066	0.0810	0.0433
SVR	0.8960	0.0278	0.1261	0.0645	0.8446	0.0278	0.1668	0.0921

The result from table 2 clearly shows that out of three ML models random forest performs with greater accuracy and outperform ANN as well as SVR in predicting adsorption capacity. RF achieving not only the highest training accuracy (R^2 = 0.9953) but also the strongest predictive capability (R^2 = 0.9700; Fig. 2a&b). The very low training errors indicate that RF effectively captures complex nonlinear relationships while also generalizing it. The testing performance looks satisfactory, with minimal RMSE.

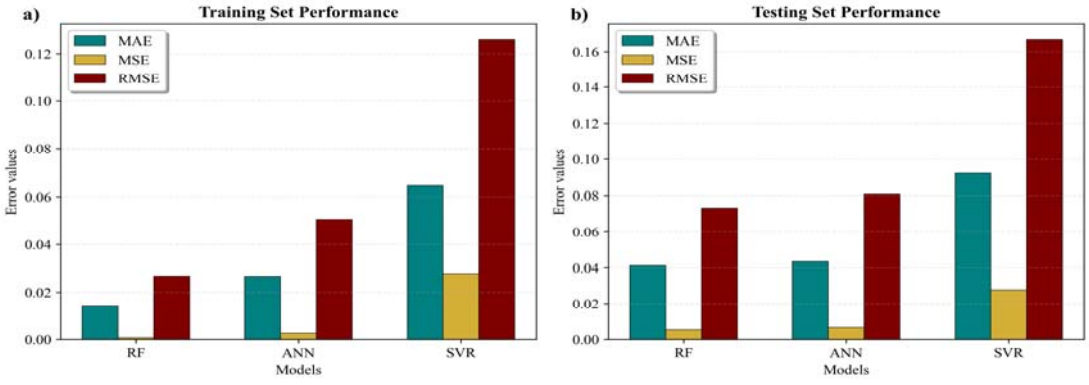


Fig. 1 Graphical representation of error values of ML models.

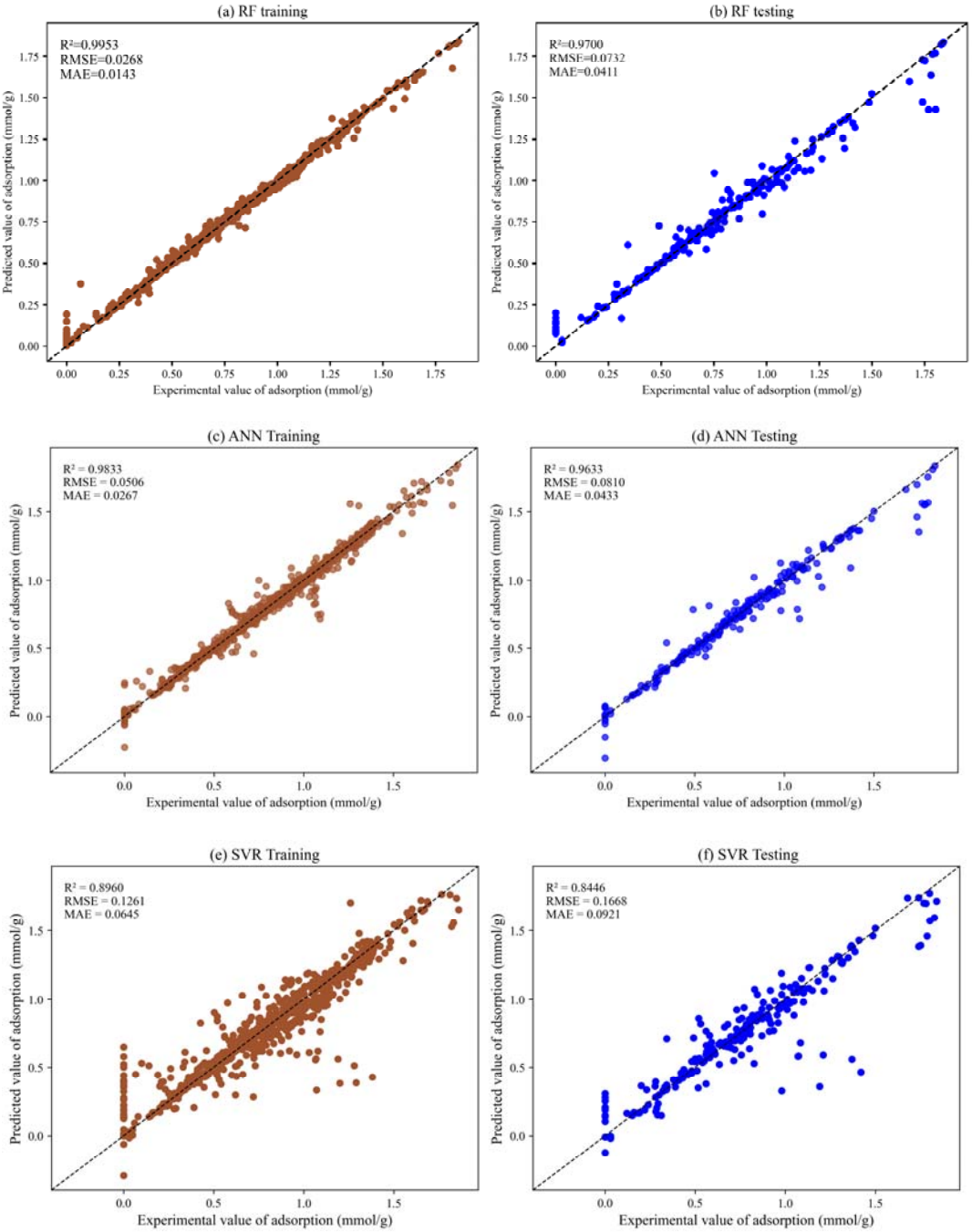


Fig. 2 Training and testing performance of (a-b) RF, (c-d) ANN, and (e f) SVR model.

3.2 Validation of the Model

Four independent coal samples (Figure 3) with contrasting physicochemical properties were selected for external validation of the developed model. The selected samples cover moisture values from 0.60% to 1.40%, volatile matter from 23.3% to 52.9%, ash yield from 2.9% to 20.0%, fixed carbon from 67.8% to 85.4%, VRr from 1.00% to 2.41%, TOC from 73.1 wt.% to 85.3 wt.%, and temperature from 19.42 °C to 55.05 °C. This broad variation in coal rank, composition, and testing conditions enables assessment of model sensitivity under diverse scenarios. Such validation supports evaluation of the predictive accuracy, robustness, and generalization ability of the model for CO_2 adsorption estimation in coal.

The RF model was evaluated against these four completely unseen validation samples, and the resulting performance metrics tells that the model achieved R^2 values of 0.9971, 0.9658, 0.9883, and 0.9814 for Samples 1 through 4, respectively, with corresponding RMSE values of 0.0201, 0.0370, 0.0234, and 0.0139 mmol/g. All four samples show an R^2value greater than 0.965 confirming consistent and reliable predictive accuracy across different coal types and experimental conditions.

The highest accuracy was achieved for Sample 1 (R^2 = 0.9971 , RMSE = 0.0201 mmol/g) indicating the model's excellent ability to capture adsorption behaviour in high rank coal with high TOC (83.2 wt.%) and vitrinite reflectance (VRr = 2.41%).Sample 4, despite having the lowest coal rank among the validation set (VRr = 1.00%), achieved the lowest RMSE of 0.0139 mmol/g, indicating excellent point-wise prediction accuracy. Sample 2 recorded the relatively lower R^2 (0.9658) and highest RMSE (0.0370 mmol/g) among the four, which may be attributed to its non-monotonic adsorption behaviour at elevated pressures - a supercritical CO_2 effect commonly reported in the literature (Gensterblum et al., 2010; Weniger et al., 2012) -where adsorption capacity initially increases with pressure before declining, a trend that is inherently more challenging to predict precisely.

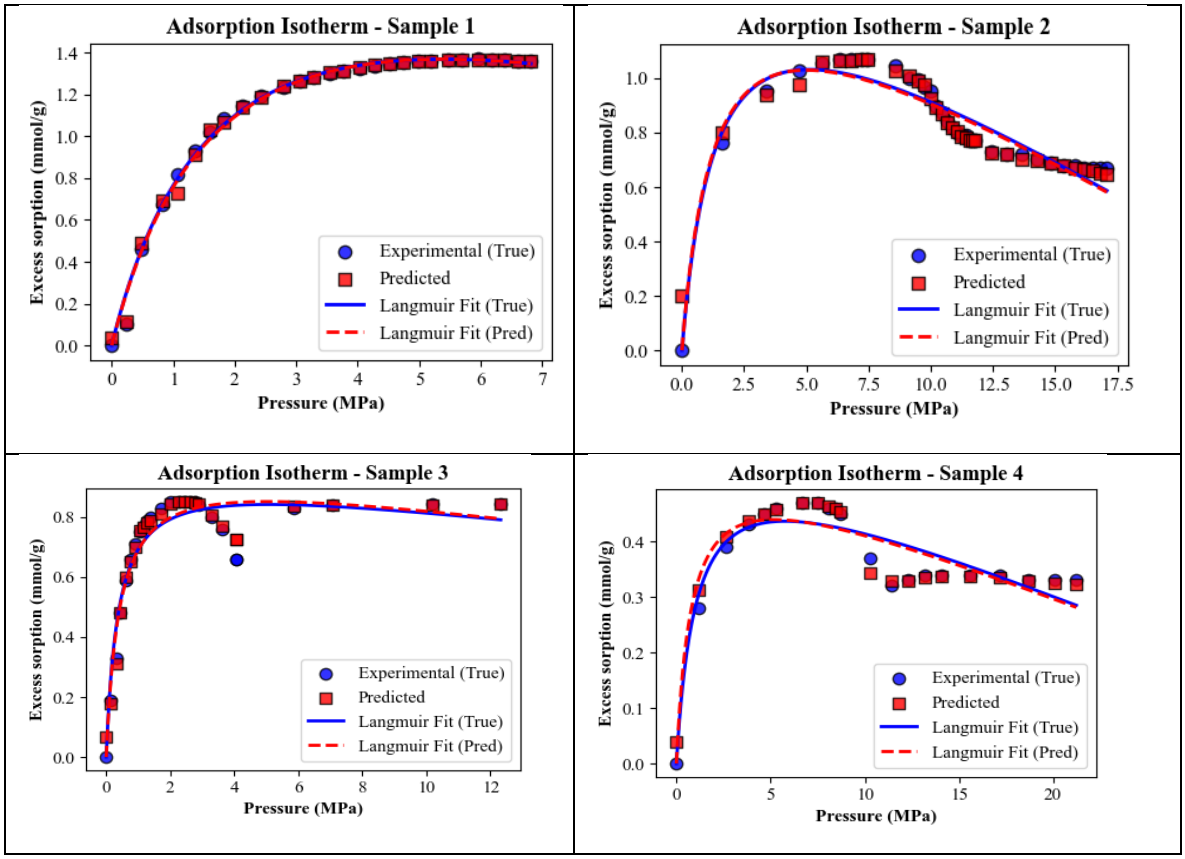


Fig. 3 Experimental and predicted Langmuir adsorption isotherms for (a) Sample 1, (b) Sample 2, (c) Sample 3 and (d) Sample 4.

3.3 Geological Interpretation based on SHAP Analysis

To evaluate the effect of individual input variables on CO_2 adsorption capacity, a sensitivity analysis was carried out using the SHAP (Shapley Additive exPlanations) method. The SHAP summary plot (Figure 4) indicates that equilibrium pressure is the most influential parameter. Increasing pressure generally results in increment of adsorption until saturation, after which negative contributions appear, indicating reduced adsorption at very high pressures, consistent with findings reported by Gensterblum et al. (2010) and Weniger et al. (2012).

Moisture is the second most important factor and mostly shows a negative effect, where higher moisture lowers adsorption capacity. Volatile matter shows an inverse relationship, supporting the trend of higher adsorption with increasing coal rank. Temperature has only a weak negative influence at higher values. Other parameters such as VRr, TOC, ash and fixed carbon contribute very little individually and the SHAP values are centered around zero. The overall SHAP ranking is consistent with the importance of features in the model, which supports the reliability and interpretability of the model.

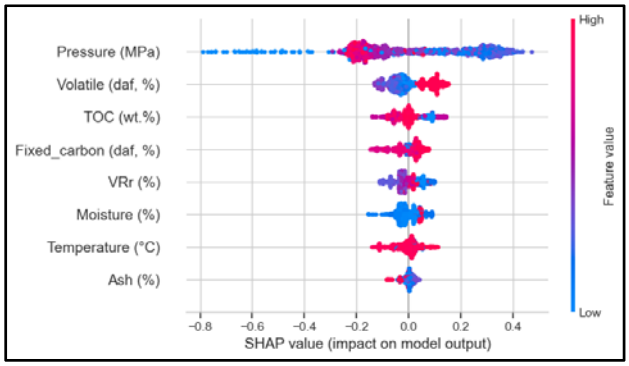


Fig.4 SHAP Summary Plot Showing the Influence of Input Parameters on CO_2 Adsorption Capacity Prediction.

4. Conclusion

The most important factor governing CO_2 adsorption is reservoir pressure, controlling rapid low-pressure uptake, intermediate saturation and post-saturation decline. The RF model captured this non-linear behaviour with good agreement with the "experimental data R^2 above 0.97". Pressure, moisture, volatile matter and TOC were identified as dominant controls through SHAP analysis.

The small difference between training and testing accuracy suggests generalization with limited overfitting. This framework provides a rapid and cost-effective tool for preliminary ECBM-CCUS reservoir screening. Future work should incorporate pore-scale descriptors such as micropore volume and BET surface area, alongside probabilistic uncertainty quantification, to further strengthen the framework's applicability at field scale.

5. References

Ahmed, R., Liu, G., Yousaf, B., Abbas, Q., Ullah, H., Ali, M.U., 2020. Recent advances in carbon-based renewable adsorbent for selective carbon dioxide capture and separation-A review. J Clean Prod 242, 118409.

Gensterblum, Y.; van Hemert, P.; Billemont, P.; Battistutta, E.; Busch, A.; Krooss, B. M.; De Weireld, G.; Wolf, K.-H. A. A. European Inter-Laboratory Comparison of High Pressure CO2 Sorption Isotherms II: Natural Coals. *Int. J. Coal Geol.***2010**, *84* (2), 115–124.

Jiang, L.; Chen, Z.; Farouq Ali, S. M.; Zhang, J.; Chen, Y.; Chen, S. Storing Carbon Dioxide in Deep Unmineable Coal Seams for Centuries Following Underground Coal Gasification. *J. Clean. Prod.***2022**, *378*, 134565.

Langmuir, I. THE ADSORPTION OF GASES ON PLANE SURFACES OF GLASS, MICA AND PLATINUM. *J. Am. Chem. Soc.***1918**, *40* (9), 1361–1403.

Tavakolian, M., Najafi-Silab, R., Chen, N. and Kantzas, A., 2024. Modeling of methane and carbon dioxide sorption capacity in tight reservoirs using Machine learning techniques. *Fuel*, *360*, p.130578.

Weniger, P.; Franců, J.; Hemza, P.; Krooss, B. M. Investigations on the Methane and Carbon Dioxide Sorption Capacity of Coals from the SW Upper Silesian Coal Basin, Czech Republic. *Int. J. Coal Geol.***2012**, *93*, 23–39.