Research on Prediction of Mudstone Breakthrough Pressure Based on Support Vector Machine in CO_2 Geological Storage

Jianhong Lin[1], Yongjie Ma[2,3,4,*], Yu Zhang[5]

[1]China Water Northeastern Investigation, Design and Research Co., Ltd., Changchun, 130021, China
[2]Zhejiang Huadong Geotechnical Investigation & Design Institute CO., LTD, Hangzhou, China
[3]PowerChina Huadong Engineering Corporation Limited, Hangzhou, China
[4]Faculty of Engineering, China University of Geosciences, Wuhan, China
[5]School of Mechanics and Civil Engineering, China University of Mining and Technology, Xuzhou, China

Abstract. This study aims to use the Support Vector Machine (SVM) model to predict the breakthrough pressure of mudstone. By collecting data on porosity, permeability, specific surface area, and maximum throat radius from 55 sets of mudstone samples, using them as input factors and breakthrough pressure of mudstone as output factors, an SVM model was constructed and trained. The research results show that the established SVM model has high prediction accuracy and good generalization ability, and can accurately predict the breakthrough pressure of mudstone. Grid search and analysis of the penalty parameter C and kernel parameter γ in the SVM model revealed the existence of specific optimal parameter combinations that can improve model performance. This study provides an effective method for predicting the breakthrough pressure of mudstone, and also provides a scientific basis for a deeper understanding of mudstone permeability and its application in CO_2 geological storage.

1 Introduction

The amount of CO_2 emissions in the atmosphere is increasing year by year. How to control and reduce the carbon dioxide content is currently a hot research topic in the world and a technical challenge [1]. CO_2 geological storage refers to the use of deep, low-permeability underground spaces to store CO_2. Carbon dioxide geological storage is considered a safe, reliable, and low-risk CO_2 reduction method for effectively treating excess gas and reducing carbon dioxide content [2]. In CO_2 geological storage, the breakthrough pressure of cap rock is one of the key parameters for evaluating its sealing performance [3]. The breakthrough pressure of mudstone, which is the minimum pressure required for fluid to pass through mudstone, is an important indicator for evaluating its permeability and sealing performance [4]. Accurately predicting the breakthrough pressure of mudstone is of great significance for evaluating the sealing ability of the cap rock and guiding oil and gas exploration and development.

Scholars at home and abroad have achieved certain results in the study of pressure breakthrough in mudstone. In the early days, the research mainly focused on the relationship between the macroscopic characteristics of mudstone cap rocks and breakthrough pressure. With the development of technology, it gradually deepened into the study of the influence of microstructure on breakthrough pressure [5]. On the one hand, using experimental testing techniques, breakthrough pressure tests were conducted on mudstone samples from different regions to analyze their influencing factors [6]. Through extensive core experiments, the correlation between physical parameters such as porosity, permeability, and specific surface area of mudstone and breakthrough pressure has been studied [7]. On the other hand, research on predicting shale breakthrough pressure based on logging data is being conducted. By establishing a mathematical model of logging parameters and breakthrough pressure, continuous calculation and regional prediction of shale breakthrough pressure can be achieved [8]. However, due to the complexity and heterogeneity of mudstone structure, predicting breakthrough pressure has always been a challenge. Traditional prediction methods often rely on experimental testing, which is not only time-consuming and laborious, but also difficult to fully reflect the permeability characteristics of mudstone.

With the rapid development of computer technology and machine learning algorithms, using data-driven methods for predicting shale breakthrough pressure has become a hot research topic. Support Vector Machine (SVM), as a powerful machine learning algorithm, has shown unique advantages in small sample, nonlinear, and high-dimensional pattern recognition fields [9]. SVM has achieved good application results in many fields, such as image recognition, bioinformatics, fault diagnosis, etc. [10]. Applying SVM to predict breakthrough pressure in mudstone is expected to overcome the shortcomings of traditional methods and improve the accuracy and efficiency of prediction.

* Corresponding author: yongjiema@foxmail.com

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

Therefore, this article aims to construct a mudstone breakthrough pressure prediction model based on support vector machine, achieve accurate prediction of mudstone breakthrough pressure, and optimize the model parameters, in order to provide a new idea and method for evaluating mudstone permeability, and help solve mudstone related problems in CO_2 geological storage, oil and gas exploration, underground engineering and other fields.

2 Methodology

The workflow used in this study is detailed in Figure 1, and each step will be described one by one below.

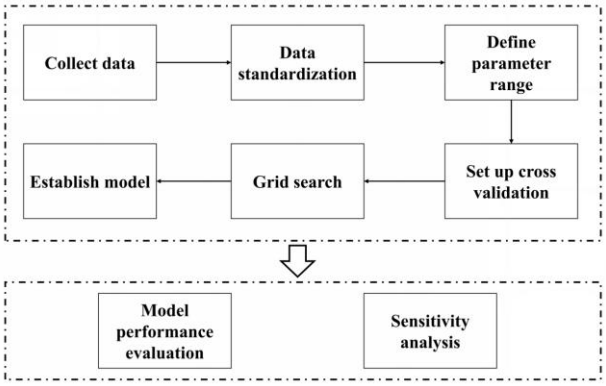

Figure 1. Schematic diagram of workflow for application research of support vector machine model.

2.1 Data collection and characterization

This study collected data on 55 sets of mudstone samples from the research dataset and relevant literature of the research group [11] [12] [13], including porosity, permeability, specific surface area, maximum throat radius, and breakthrough pressure. Porosity refers to the ratio of pore volume to total volume in mudstone, which reflects the degree of pore development in mudstone and has a significant impact on fluid storage and transport. Permeability represents the ability of mudstone to allow fluid to pass through, and is a key indicator for measuring the permeability of mudstone. The specific surface area is the surface area per unit mass of mudstone, which affects the interaction between mudstone and fluids. The maximum roaring radius determines the size of the largest connected pores in mudstone and has a direct impact on breakthrough pressure. Table 1 provides statistical information of the dataset, including mean, maximum, minimum, and standard deviation.

Table 1. Schematic diagram of workflow for application research of support vector machine model.

Parameter	Average value	Maximum value	Minimum value	Standard deviation
Porosity (%)	7.99	29.9	0.06	9.22
Permeability (nD, $10^{-21}m^2$)	209.85	4325	0.08	662.42
Specific surface area (m^2/g)	46.8	1.47	14.96	11.79
Maximum roar radius (nm)	22.51	3.00	260.00	37.34
Breakthrough pressure(MPa)	6.30	18.8	0.12	5.51

2.2 Principle of Support Vector Machine Model

Originally developed for binary classification tasks, the Support Vector Machine (SVM) operates by constructing an optimal separating hyperplane within high-dimensional feature spaces. This method aims to maximize the margin between data points representing distinct classes, thereby enhancing generalization capabilities and ensuring robust classification performance. The fundamental principle underlying SVM involves identifying a decision boundary that optimally segregates categorical instances while maintaining maximal separation distances in transformed feature representations [9]. In regression problems, support vector regression introduces slack variables and penalty parameters to allow the model to have some degree of error, thereby finding an optimal regression function. For a given training dataset, the goal of SVR is to find a function f(x) that minimizes training errors while maintaining low model complexity. Kernel functions frequently employed in machine learning models include linear kernels, polynomial kernels, and radial basis functions (RBF), among others. Among these, radial basis functions are predominantly utilized in support vector regression (SVR) owing to their superior local adaptability and strong generalization capabilities. These characteristics enable RBF kernels to effectively capture complex patterns in data while maintaining robust performance across diverse applications [9].

2.3 Establishment and Training of Support Vector Machine Model

2.3.1 Data standardization

Firstly, the Z-score function is used to standardize the input features, unifying the data of different features to the same scale to avoid bias in model training caused by significant differences in feature scales.

2.3.2 Define parameter range

The penalty parameter C controls the tolerance of the model for errors and determines the trade-off between training errors and model complexity during the training process [10]. The kernel function parameter γ determines the width of the radial basis function, which affects the distribution of data in high-dimensional space [10]. In this study, the range of penalty parameter C was set to logspace (-2,2,10), and the range of kernel function parameter γ was set to logspace (-2,2,10). By searching for the optimal parameters within this range, the best model performance was obtained.

2.3.3 5-fold cross validation

5-fold cross validation is the process of randomly dividing a dataset into 5 disjoint subsets, selecting 4 subsets as the training set and the remaining 1 subset as the testing set for 5 training and testing sessions. On each fold, train the support vector machine model using the training set and test it using the test set to calculate the root mean square error (RMSE) on the test set. Through 5-fold cross validation, the performance of the model on different subsets of data can be more comprehensively evaluated, avoiding evaluation bias caused by the randomness of data partitioning.

2.3.4 Grid Search

Traverse all possible parameter combinations, i.e. traverse the range of values for penalty parameter C and kernel function parameter γ, perform 5-fold cross validation for each parameter combination, and calculate the average RMSE. Select the parameter combination with the minimum average RMSE as the optimal parameter. Through this method, the optimal penalty parameter C and kernel function parameter γ of the support vector machine model were determined, and a shale breakthrough pressure prediction model based on support vector machine was established.

3 Results and discussion

3.1 Model performance evaluation

To assess the developed support vector machine (SVM) model's capacity for forecasting mudstone breakthrough pressure, we partitioned the dataset into 70% training and 30% testing subsets. Following model training on the designated training portion, predictions were generated for the testing subset. Performance evaluation involved computing R^2, mean absolute error (MAE), and root mean square error (RMSE) metrics for both subsets. R^2 serves as an indicator of model-data alignment, with values nearing 1 signifying superior fit quality. MAE quantifies the average magnitude of discrepancies between predicted and observed values, providing insight into typical prediction accuracy. RMSE, by assigning greater emphasis to substantial errors, offers a more sensitive measure of model precision, effectively highlighting cases where predictions significantly diverge from actual measurements. This comprehensive metric set enables robust assessment of both the model's generalization capability and its reliability in practical applications.

As shown in Figure 2, the R^2 on the training set reached 0.9707, MAE was 0.6140, and RMSE was 0.8814, indicating that the model has a good fitting effect on the training set and can accurately capture the relationship between input factors and shale breakthrough pressure. As shown in Figure 3, the R^2 on the test set is 0.9798, MAE is 0.6763, and RMSE is 0.8753, indicating that the model has good generalization ability and can accurately predict the breakthrough pressure of unseen mudstone samples to a certain extent. The R^2 of all data is 0.9741, MAE is 0.6321, and RMSE is 0.8797 (Figure 4). Compared with other traditional prediction methods, the support vector machine model constructed in this study performs better in terms of prediction accuracy and generalization ability. Traditional empirical formula methods are often based on specific geological conditions and experimental data, with limited applicability and large prediction errors in complex geological conditions. The support vector machine model can automatically learn complex patterns and relationships in data, without being limited by empirical formulas, and can better adapt to predicting shale breakthrough pressure under different geological conditions.

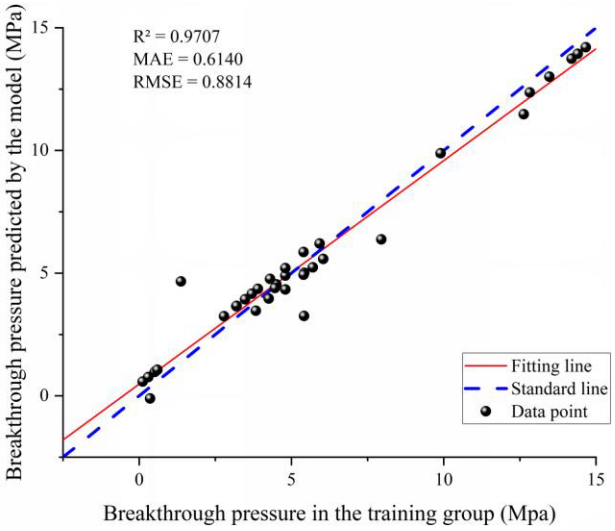

Figure 2. Training group SVM model prediction results.

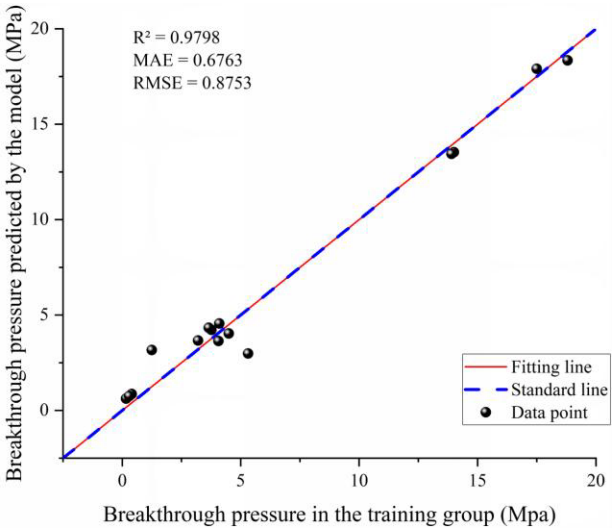

Figure 3. Test group SVM model prediction results.

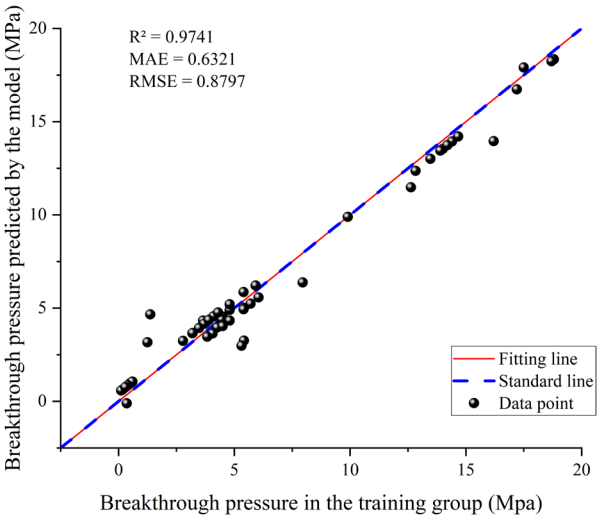

Figure 4. All data SVM model prediction results.

3.2 Sensitivity analysis of factors affecting shale breakthrough pressure

In the construction of Support Vector Machine (SVM) models, the penalty parameter C and kernel parameter γ are crucial hyperparameters whose values directly affect the performance of the model. In order to thoroughly analyze the impact of different combinations of C and γ on root mean square error (RMSE), this study conducted a detailed parameter grid search and plotted the corresponding parameter grid search heatmap, as shown in Figure 5. It can be clearly observed from the heatmap that there is a complex nonlinear relationship between the values of C and γ and RMSE. In the horizontal direction, as the C value gradually increases from a small range, RMSE initially shows a downward trend, indicating that moderately increasing the C value helps the model better fit the training data and reduce prediction errors. This is because the C value controls the punishment level of the model for misclassified samples, and a larger C value will make the model more demanding on the fitting accuracy of the training data. However, when the C value exceeds a certain threshold, RMSE begins to rise, which means that the model has overfitting and its generalization ability to new data is weakened. In the vertical axis direction, the change in gamma value also has a significant impact on RMSE. A smaller gamma value allows the radial basis function to have a wider range of action, and the decision boundary of the model is relatively smooth. At this time, the RMSE may be larger because the model has weaker ability to capture local features of the data. As the gamma value increases, the model becomes more sensitive to local features of the data and can better fit the training data, resulting in a decrease in RMSE. But when the gamma value is too high, the model is prone to overfitting the noise and outliers in the training data, leading to a further increase in RMSE. Through a comprehensive analysis of the heat map, there exists a relatively optimal C-γ combination (C=100, γ=0.5995) with a relatively low RMSE value, indicating that the model has good predictive performance under this parameter combination. The grid search process in this study is aimed at accurately finding the optimal parameter combination to improve the accuracy and reliability of the model in predicting shale breakthrough pressure.

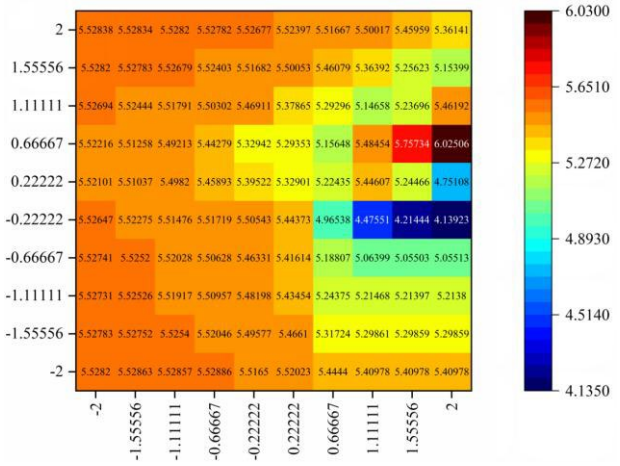

Figure 5. Support Vector Machine Model Superparameter Grid Search Heat Map.

4 Conclusions

This investigation developed a support vector machine (SVM) framework for estimating shale breakthrough pressure through analysis of 55 shale samples, the following main research results were achieved:

(1) The SVM-based predictive model demonstrates exceptional performance, evidenced by elevated R^2 coefficients and minimized mean absolute error (MAE) and root mean square error (RMSE) across both calibration and validation datasets. These metrics confirm the model's precision and strong extrapolation capacity for accurate shale pressure estimation.

(2) Mudstone characteristics—specifically porosity, permeability, specific surface area, and maximum throat radius—were identified as salient predictors for breakthrough pressure estimation, establishing a robust foundation for SVM input selection.

(3) Sensitivity analysis revealed that the regularization parameter (C) and radial basis function spread (γ) critically govern model behavior. Systematic grid search optimization yielded an effective parameter configuration (C=100, γ=0.5995), under which the model achieves optimal balance between complexity and generalization, ensuring reliable predictive outcomes.

Acknowledgment

The authors gratefully acknowledge the supports of the Natural Science Foundation of Jiangsu Province (No. BK20231080) and the National Natural Science Foundation of China (No.42302273), China Postdoctoral Science Foundation (No.2022M713367).

References

1. Yanwei Wang, Hongyang Chu, Xiaocong Lyu. Deep learning in CO_2 geological utilization and storage: Recent advances and perspectives. [J]. Advances in Geo-Energy Research, 2024, Vol.13(3): 161-165.
2. Linlin ZHANG, Fengpeng LAI, Yintao DONG, Yuting DAI. Evaluation of CO_2 storage effected by geological parameters of brine layer[J]. Meitan xuebao, 2024, Vol.49(9): 3932-3943.
3. Jianwen, D.; Yukun, C.; Yuting, J.; Li, L.; Shenghao, W.; Lei, S.; Jian, T.; Chaozhong, Q.; Qiang, C.; Quan, G. Limiting Pathways and Breakthrough Pressure for CO_2 Flow in Mudstones. The National Key Laboratory of Water Disaster Prevention, Nanjing 210029, China;State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource and Hydropower, Sichuan University, Chengdu 610065, China;Y 2023, Vol.625, 129998, doi:10.1016/j.jhydrol.2023.129998.
4. Chen, B.; Li, Q.; Tan, Y.; Yu, T.; Li, X.; Li, X. Experimental measurements and characterization models of caprock breakthrough pressure for CO_2 geological storage. State Key Laboratory of Geomechanics and Geotechnical Engineering, Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, Wuhan 430071, China; University of Chinese Academy of Sciences, Beijing 100049, China 2024, Vol.252, 104732, doi:10.1016/j.earscirev.2024.104732.
5. Kim, S.-O.; Wang, S.; Lee, M. Evaluation of hydrogeologic seal capacity of mudstone in the yeongil group, pohang basin, Korea: Focusing on mercury intrusion capillary pressure analysis. Department of Energy Resources Engineering, Pukyong National University, South Korea; Department of Earth Environmental Sciences, Pukyong National University, South Korea 2020, Vol.53, 23-32, doi:10.9719/eeg.2020.53.1.23.
6. Li, Y.; Zha, M.; Song, R.; Aplin, A.C.; Bowen, L.; Wang, X.; Zhang, Y. Microstructure and pore systems of shallow-buried fluvial mudstone caprocks in Zhanhua depression, east China inferred from SEM and MICP. Post-Doctoral Research Station of Geological Resource and Geological Engineering, Chengdu University of Technology, Chengdu, Sichuan, 610059, China College of Energy Resources, Chengdu University of Technology, Chengdu, Sichua 2021, Vol.132, 105189, doi:10.1016/j.marpetgeo.2021.105189.
7. Hao, S.; Cao, J.; Jiang, Y.; Zhang, S.; Zhou, L. Analysis of rock microstructure after CO_2 breakthrough based on CT scanning. ; [1] China Univ Geosci Wuhan, Sch Environm Studies, Fac Engn, Wuhan 430074, Peoples R China [2] Three Gorges Co Ltd, Survey Res Inst, Wuhan 430074, Peoples R China [3] Univ Erlangen Nurnberg, Geoctr Northern Bavaria, S 2024, Vol.83, 1-11, doi:10.1007/s12665-024-11750-8.
8. Gao, Y.; Ren, Z.; Chen, M.; Jiang, H.; Ding, S. Coupled geomechanical-thermal simulation for oil sand reservoirs with shale barriers under hot water injection in vertical well-assisted SAGD wells. Geothermal Research Institute of the Dagestan Scientific Center of the Russian Academy of Sciences, 367003 Makhachkala, Shamilya Str. 39-A, Dagestan, Russia 2022, Vol.208, 109644, doi:10.1016/j.petrol.2021.109644.
9. Xiaoyu Li, S.D., Shaoyang Guo, Chanjin Zheng. Applying support vector machines to a diagnostic classification model for polytomous attributes in small-sample contexts. Affiliations Lab of Artificial Intelligence for Education, East China Normal University, Shanghai, China. School of Computer Science and Technology, East China Normal University, Shanghai, C 2024, doi:10.1111/bmsp.12359.
10. Liu, Y.; Gu, H.; Qin, P. A verified training support vector machine in bearing fault diagnosis. School of Control Science and Engineering, Dalian University of Technology, 2 Linggong Road, Dalian 116024, Liaoning, People's Republic of China 2024, Vol.35, 116131, doi:10.1088/1361-6501/ad6c75.
11. Hildenbrand, A.; Schlömer, S.; Krooss, B.M.; Littke, R. Gas breakthrough experiments on pelitic rocks: comparative study with N2, CO_2 and CH4. 1 Lehrstuhl für Geologie, Geochemie und Lagerstätten des Erdöls und der Kohle, Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen, Lochnerstr. 4-20, D-52056-Aachen, Germany; 2 EniTechnologie SPA, Via F. Maritano 26, San Donato Milane 2004, vol.4, 61-80, doi:10.1111/j.1468-8123.2004.00073.x.
12. Shuren Hao. Microscopic Characteristics and Modeling of CO, Traps in Mudstone Caprocks, Jilin University, 2019.
13. Alexandra Amann-Hildenbrand, P.B., Andreas Busch, Bernhard M. Krooss. Experimental investigation of the sealing capacity of generic clay-rich caprocks. Institute of Geology and Geochemistry of Petroleum and Coal, RWTH Aachen University, Germany Clay and Interface Mineralogy, RWTH Aachen University, Germany Shell Global Solutions International B.V., The Ne 2013, Vol.19, 620-641, doi:10.1016/j.ijggc.2013.01.040.