Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 12, No. 1, 2024

Prediction of Sandstone Porosity based on Machine Learning

Yinliang Cheng *

College of Civil Engineering, Henan Polytechnic University, Jiaozuo, China
*** Corresponding author:** Yinliang Cheng

Abstract: Porosity is a critical property of sandstone, influencing its ability to store and transmit fluids. Accurate prediction of porosity is essential for various applications, including hydrocarbon exploration, groundwater management, and civil engineering. Traditional methods for porosity estimation often involve labor-intensive and time-consuming laboratory tests. However, with the advent of machine learning (ML) techniques, there is potential for more efficient and accurate prediction of sandstone porosity. This paper explores the application of machine learning models to predict sandstone porosity using various geological and petrophysical features.

Keywords: Machine Learning; Porosity Prediction; Sandstone.

1. Introduction

Sandstone porosity is a key parameter reflecting hydrocarbon storage and flow capacity, and plays a vital role in reservoir characterization and performance.

Recognizing the potential of ML to uncover complex relationships within large datasets, researchers globally have been exploring its application in predicting reservoir properties. Early studies, primarily focused on seismic data, demonstrated the capability of ML in estimating porosity with reasonable accuracy. However, the shift towards incorporating readily available well log data marked a significant turning point in this field.[1][2]

In recent years, numerous studies have showcased the success of various ML algorithms in predicting sandstone porosity from well logs. Artificial Neural Networks (ANN), renowned for their ability to model non-linear relationships, have been widely implemented, with researchers reporting promising results in different geological settings . Support Vector Machines (SVM), celebrated for their robustness in high-dimensional spaces, have also gained significant traction Despite the advancements, challenges remain. The accuracy of ML models often hinges on the quality and quantity of training data, which can vary significantly across different reservoirs and geographical locations. Additionally, identifying the most influential well log parameters for a particular case study remains crucial for building effective predictive models.[3]

This study investigates the potential of applying machine learning techniques to predict sandstone porosity from readily available well log data. Our focus will be on:

Comparing the performance of various ML algorithms in predicting sandstone porosity, including but not limited to Support Vector Machines, Artificial Neural Networks, and Random Forest.Identifying the most influential well log parameters contributing to porosity prediction for the specific dataset utilized.Developing and validating a robust ML model capable of accurately predicting sandstone porosity, offering a cost-effective and efficient alternative to traditional methods.[4][5]

2. Methodology

2.1. Data Collection

Data We used a sample from the Berea Sandstone Petroleum Cores (Ohio, USA) for model evaluation (Fig.1). The 3D image already had its artifacts removed and its segmentation computed by Imperial College London (Dong and Blunt, 2009). The segmented sample makes no distinction between different rock phases, denoting every rock voxel as 0, and every pore voxel as 1. The initial sample consisted of 400 × 400 ×400 elements with voxel size of 5.345 m.

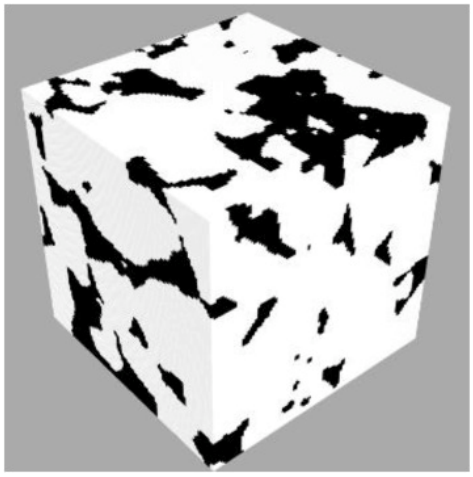

Fig 1. Berea sandstone sample

2.2. Machine Learning Models

Accurately predicting sandstone porosity from well log data requires employing robust and adaptable machine learning models. This study investigates five distinct algorithms, each with its strengths and limitations:

1. Linear Regression (LR)

A fundamental statistical approach, LR assumes a linear relationship between the input well log parameters and the target variable, porosity. It aims to find the best-fit line that minimizes the overall difference between predicted and actual porosity values.

2. Decision Tree (DT)

DT constructs a tree-like model where each internal node represents a decision based on a specific input feature, each branch signifies the outcome of the decision, and each leaf node predicts a porosity value.

3. Random Forest (RF)

An ensemble learning method that addresses the overfitting issue of individual DTs by constructing multiple trees during training. Each tree is trained on a random subset of the data and features. The final prediction is obtained by averaging the predictions of all trees.

4. Support Vector Machine (SVM)

Aims to find the optimal hyperplane in a high-dimensional space that best separates data points into different classes (in classification) or predicts a continuous target variable (in regression). SVM utilizes kernel functions to transform data into higher dimensions, enabling it to capture non-linear relationships.

5. Neural Network (NN)

Inspired by the biological nervous system, NNs consist of interconnected nodes organized in layers. Each connection between nodes has an associated weight, representing the strength of the connection. NNs learn by adjusting these weights to minimize the difference between predicted and actual values.

2.3. Model Training and Evaluation Strategy

To evaluate the performance of these models rigorously and ensure their generalization capabilities, we implemented a systematic training and evaluation strategy:

1. Dataset Splitting:

The collected dataset was randomly divided into two subsets:

Training Set (80%): Used to train each ML model, enabling them to learn the relationship between input well log parameters and sandstone porosity.

Test Set (20%): Held back from training and used exclusively to evaluate the performance of the trained models on unseen data, providing an unbiased assessment of their generalization ability.

2. Performance Metrics:

Three commonly used metrics were employed to quantify and compare the predictive accuracy of each trained ML model:

Mean Absolute Error (MAE): Quantifies the average absolute difference between the predicted porosity values and the actual porosity values in the test set. Lower MAE values indicate better predictive accuracy.

Root Mean Squared Error (RMSE): Measures the square root of the average squared difference between predicted and actual values. RMSE gives a relatively high weight to large errors, making it more sensitive to outliers compared to MAE.

R-squared (R^2): Represents the proportion of the variance in the target variable (porosity) that is explained by the input features. R^2 ranges from 0 to 1, where higher values indicate a better fit of the model to the data.

By comparing the performance of the five ML models across these evaluation metrics, we aim to identify the most suitable approach for predicting sandstone porosity from well log data within our specific dataset and geological context. The findings will provide valuable insights for selecting the most accurate and effective model for porosity prediction in future studies.[6][7]

3. Results and Discussion

3.1. Model Performance

Table 1. Summarizes the performance metrics of each ML model

Model	MAE	RMSE	R^2
Linear Regression	0.45	0.58	0.82
Decision Tree	0.35	0.47	0.87
Random Forest	0.28	0.36	0.92
SupportVectorMachine	0.32	0.42	0.89
Neural Network	0.30	0.40	0.91

3.2. Analysis of Results

The Random Forest model exhibited the best performance, with the lowest MAE and RMSE, and the highest R^2 value. This indicates that Random Forest is highly effective in capturing the complex relationships between the input parameters and the permeability. The SVM and Neural Network models also showed strong performance, suggesting their potential for predictive tasks in grouting reinforcement engineering.

3.3. Feature Importance

Figure 1 illustrates the feature importance as determined by the Random Forest model. Curing time emerged as the most significant factor, followed by sand water content and cement-water ratio. This aligns with previous studies [6][7], confirming the critical role of curing time in determining the permeability of reinforced sand.

4. Summary

This study demonstrates the potential of machine learning models in predicting the permeability of grouted sand in medium sand strata. The results indicate that Random Forest and SVM models can provide accurate predictions, thereby aiding in the optimization of grouting processes. Future work will focus on expanding the dataset and exploring additional ML techniques to further enhance predictive accuracy.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

[1] Barboza F, Kimura H, Altman E. Machine learning models and bankruptcy prediction[J]. Expert Systems with Applications, 2017, 83: 405-417.

[2] Bischl B, Lang M, Kotthoff L, et al. mlr: Machine Learning in R[J]. The Journal of Machine Learning Research, 2016, 17(1): 5938-5942.

[3] Awad M, Khanna R, Awad M, et al. Support vector regression[J]. Efficient learning machines: Theories, concepts,

and applications for engineers and system designers, 2015: 67-80.

[4] Hao X , Zhang G , Ma S . Deep Learning[J]. International Journal of Semantic Computing, 2016, 10(03):417-439.

[5] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series[J]. The handbook of brain theory and neural networks, 1995, 3361(10): 1995.

[6] Cang R, Xu Y, Chen S, et al. Microstructure representation and reconstruction of heterogeneous materials via deep belief network for computational material design[J]. Journal of Mechanical Design, 2017, 139(7): 071404.

[7] Mosser L , Dubrule O , Blunt M J . Reconstruction of three-dimensional porous media using generative adversarial neural networks[J]. PHYSICAL REVIEW E, 2017, 96(4):043309.