

# Reservoir Characterization and Capacity Calculation for CO<sub>2</sub> Storage Using AI/ML Techniques in Gandhar Oil Field, Cambay Basin, India

Akash Nair<sup>1</sup>, Srikumar Roy<sup>2</sup>, William K. Mohanty<sup>3</sup>, Mauricio D. Sacchi<sup>1</sup>

1. Department of Physics, University of Alberta
2. Science Foundation Ireland Research Centre in Applied Geosciences, School of Earth Sciences, University College Dublin
3. Department of Geology and Geophysics, Indian Institute of Technology Kharagpur

## 1. Summary

Geophysical investigation is essential for subsurface characterization, de-risking leakage potential, and secure CO<sub>2</sub> storage in depleted hydrocarbon reservoirs. Sustainable CO<sub>2</sub> storage in the Cambay Basin is investigated by leveraging its depleted oil and gas reservoirs. This study assesses CO<sub>2</sub> storage potential in India's Gandhar oil field in the Cambay Basin, using two approaches: i) conventional seismic and well-log data analysis and ii) machine learning-assisted reservoir characterization. Machine learning (ML) models for pay zone delineation, oil saturation prediction, and pore classification are developed by following the general ML workflow, starting from Feature Selection, Model Selection, and Model Evaluation. The models were tested with several algorithms, where Random Forest and Support Vector Machine gave the best results.

## 2. Theory / Method / Workflow

The study assesses the feasibility of carbon capture, utilization, and storage (CCUS) by evaluating key geophysical inputs for calculating CO<sub>2</sub> storage capacity. The three primary inputs considered are the thickness of the pay zone, oil saturation, and porosity within geological formations identified as suitable for CO<sub>2</sub> storage.

The first part of the study focused on using conventional reservoir characterization methods to evaluate the geophysical inputs required for calculating CO<sub>2</sub> storage. The pay zones were identified by the visual inspection of different well logs plotted against each other. The oil saturation was calculated using Archie's law, and the porosity was found using the porosity logs.

To streamline the evaluation process, the study introduces three artificial intelligence and machine learning (AI/ML) models:

1. Model 1: Automates the identification of pay zones from well-log data.
2. Model 2: Predicts oil saturation directly from well-log data, simplifying the task.
3. Model 3: Proposes predicting pore class from integrated seismic and well-log datasets.

The models follow similar steps with slight variations tailored to their specific requirements. Data preparation involves transforming raw data into a suitable format, including cleaning, transformation, normalization, and integration. The study emphasizes the importance of data quality in training accurate and reliable AI/ML models.

Feature selection, a crucial step in ML, involves choosing relevant features to build predictive models. The study uses the Pearson correlation coefficient to select features with strong correlations for further training.

Model selection involves choosing the best algorithms for each model type (Raschka, 2018). Classification algorithms are used for pay zone thickness and pore class, while regression models predict oil saturation.

The evaluation phase assesses model performance using metrics like accuracy, precision, recall, confusion matrix, and R<sup>2</sup>-score (Asuero et al., 2006). Model evaluation allows for identifying areas of improvement and adjusting model architecture or training processes.

Finally, the study discusses the calculations determining CO<sub>2</sub> storage capacity. Theoretical capacity assumes filling all available pores with CO<sub>2</sub>, while effective capacity considers factors like mobility, buoyancy, heterogeneity, water saturation, and aquifer strength. It is given by the expression:  $M_{CO_2} = \rho_{CO_2res} [R_f \times A \times h \times \phi \times (1 - S_w) - V_{iw} + V_{pw}]$  (1) where,  $\rho_{CO_2res}$  is the density of CO<sub>2</sub> at reservoir conditions, R<sub>f</sub> is the recovery factor, A is the area of the field, h is the effective thickness of the reservoir,  $\phi$  is the porosity, S<sub>w</sub> is the water saturation, V<sub>iw</sub> is the volume of injected water, and V<sub>pw</sub> is the volume of produced water. (Bachu, 2008)

The study concludes with calculations for theoretical and effective CO<sub>2</sub> storage capacities, providing a comprehensive overview of the methodology employed. In summary, the study employs AI/ML models to automate and enhance the assessment of key geophysical inputs for CO<sub>2</sub> storage capacity calculation, contributing to the advancement of CCUS technologies.

### 3. Results, Observations, Conclusions

#### 3.1 Delineation of Pay Zones from Well-Log Data

**Feature Selection:** Utilized a Pearson correlation matrix to evaluate the correlation between input parameters, highlighting relationships such as high correlation between GR and NPHI and a poor correlation of RHOB with LLD and LLS.

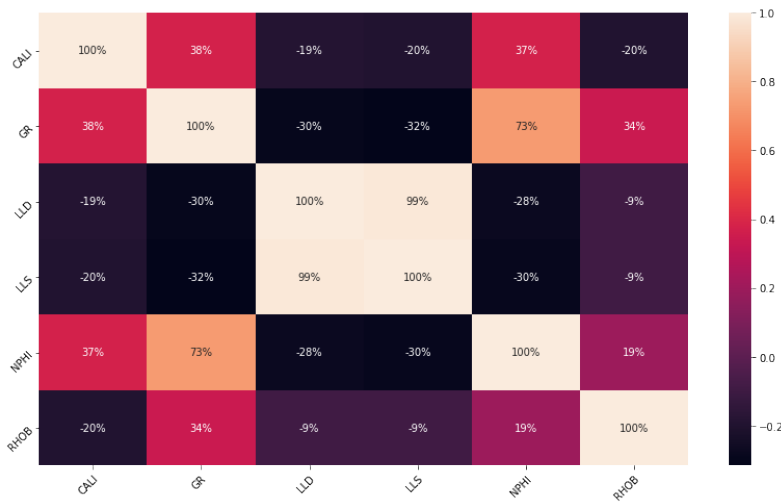


Fig 1. Pearson Correlation Matrix for Model 1

**Model Selection:** Explored different algorithms for predicting sand or shale classes, with the Random Forest Classifier outperforming others with an R2 score of 0.8583.

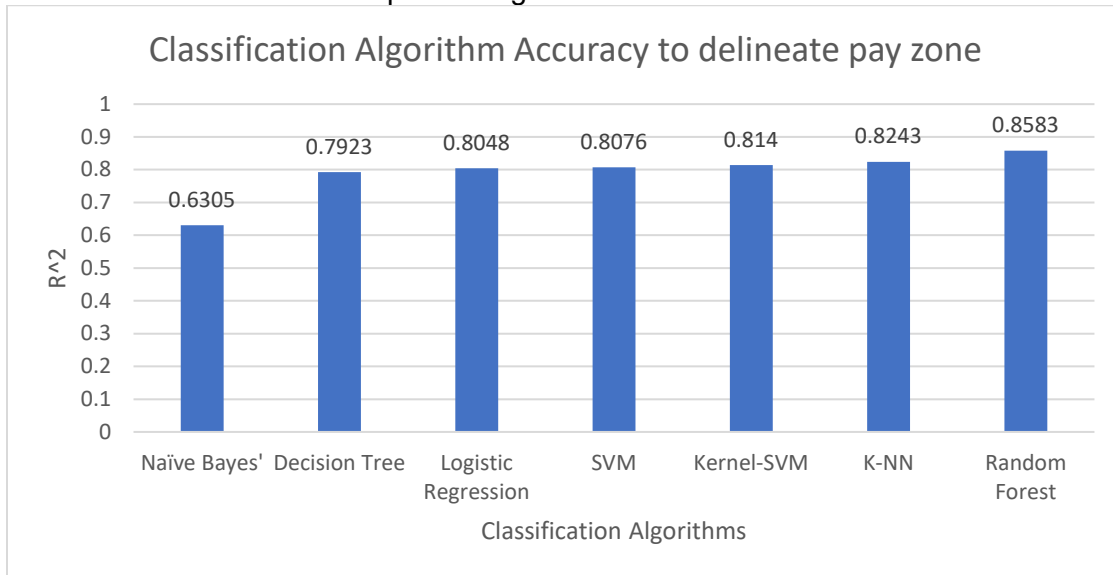


Fig 2. Comparison of different Classification algorithms for Model 1

**Model Evaluation:** Random Forest performed well in testing but faced challenges in blind well validation, particularly in accurately predicting shale due to the scarcity of shale data points during training.

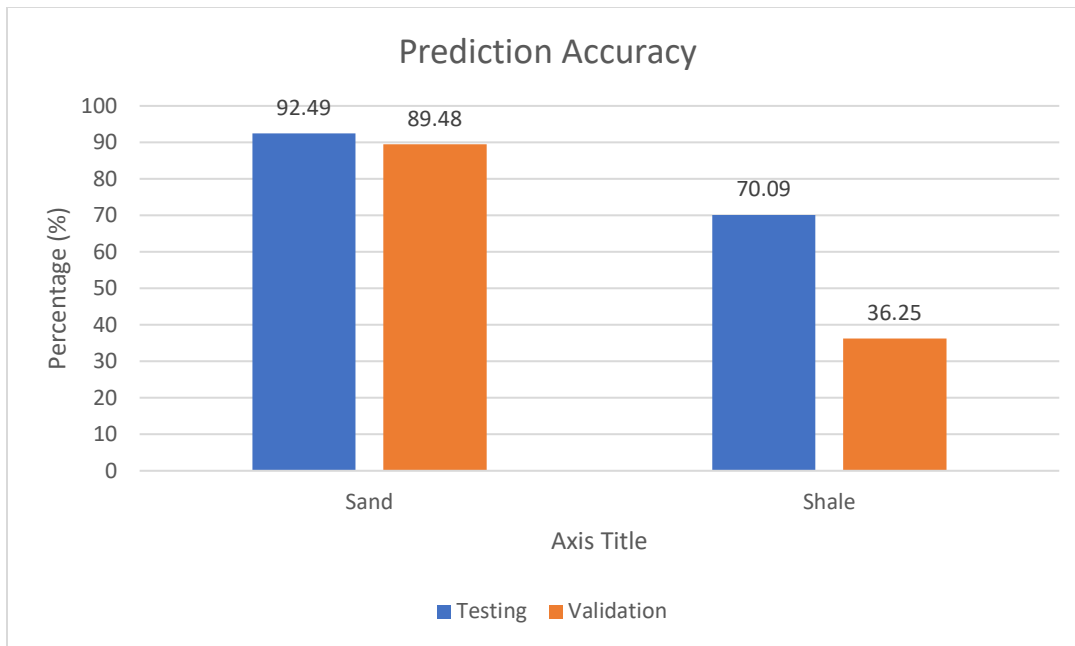


Fig 3. Accuracy of the model in sands and shales in testing and validation phases.

**Results:** Despite some discrepancies, the model worked well with the available data points, suggesting potential improvement with additional data for enhanced reservoir characterization and CO<sub>2</sub> storage evaluation.

### 3.2 Prediction of Oil Saturation from Well-Log Data

**Feature Selection:** Employed Pearson's matrix to assess correlations, selecting all inputs. Notably, SP and RHOB showed a low correlation, while RM and RT exhibited a high correlation.

**Model Selection and Evaluation:** Tested different regression algorithms, with Support Vector Regression (SVR) performing the best with an R<sup>2</sup> of 0.82 despite challenges faced by the ANN algorithm in the validation phase.

**Results:** The Oil Saturation plot illustrated good predictions during testing and validation, with SVR offering accurate predictions for high and low values.

### 3.3 Prediction of Pore Class from Integrated Seismic and Well-Log Datasets

**Feature Selection:** Pearson's matrix assessed correlations, acknowledging a non-linear relationship between porosity and seismic attributes.

**Model Selection and Evaluation:** Employed Random Forest and XGBoost due to their ability to handle non-linear relationships. Both performed well in training, but XGBoost struggled in blind well prediction.

**Results:** Achieved 54% accuracy in complete blind well prediction, indicating room for improvement in the model by incorporating more data points.

### 3.4 Calculation of CO<sub>2</sub> Storage

**Theoretical CO<sub>2</sub> Storage Capacity:** Using relevant parameters and equation 1, an estimated theoretical capacity of 11.98 MMt for the study area of approximately 50 km<sup>2</sup> is yielded.

**Effective CO<sub>2</sub> Storage Capacity:** Derived as half of the theoretical value, resulting in an estimated effective capacity of 5.99 MMt for the study area.

**Extrapolation:** The values were applied to the entire Cambay Basin, estimating an effective CO<sub>2</sub> storage capacity of 609.01 Mt, acknowledging a significant difference from a previous study and highlighting the need for further detailed studies.

This comprehensive study integrates various data analysis and modeling techniques for reservoir characterization and CO<sub>2</sub> storage assessment, recognizing areas for improvement and the importance of additional data for enhanced accuracy.

### Novel/Additive Information

This study finally highlights the ability of machine learning to develop models to evaluate geophysical inputs to identify suitable CO<sub>2</sub> injection sites and optimize injection strategies for secure CO<sub>2</sub> storage. This approach is replicable in other matured oil fields for carbon storage, considering the complexity of the depleted reservoir.

### Acknowledgments

We thank Oil and Natural Gas Corporation (ONGC) for sharing the seismic and well-log data.

### References

- Asuero, A. G., Sayago, A., & González, A. G. (2006). The correlation coefficient: An overview. In *Critical Reviews in Analytical Chemistry* (Vol. 36, Issue 1, pp. 41–59). <https://doi.org/10.1080/10408340500526766>
- Bachu, S. (2008). CO<sub>2</sub> storage in geological media: Role, means, status and barriers to deployment. In *Progress in Energy and Combustion Science* (Vol. 34, Issue 2, pp. 254–273). <https://doi.org/10.1016/j.pecs.2007.10.001>
- Raschka, S. (2018). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. <http://arxiv.org/abs/1811.12808>