Machine Learning Approach for CCS Site Screening and Selection Using Natural Analogues

Emperor Amedeka, Yen A. Sokama-Neuyam, Joshua N. Turkson, Prosper A. Akaba, Andy Ashie, Christabel B. Kusi, Edmond Ankamah; Net-Zero Carbon Emission Lab (NCEL), College of Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. Email: ncel@knust.edu.gh Tel: +233 245 9373 58 website: www.ncel.knust.edu.gh

Abstract

Carbon Capture and Storage (CCS) is a promising technology for industrial decarbonization. This study presents a machine learning framework for evaluating the long-term security of geological CO_2 storage sites using natural analogues. Three models—Random Forest, XGBoost, and Multi-Layer Perceptron—were trained on a global dataset from 76 natural CO_2 reservoirs. Random Forest outperformed other models in precision and recall. SHapley Additive exPlanations (SHAP) identified key predictors, including the presence of fault, seal integrity, and reservoir stacking. The framework provides a scalable and interpretable, data-driven solution for efficiently identifying high-potential carbon storage sites, reducing the need for costly and time-intensive preliminary assessments.

Introduction

While many aspects of the energy system can be readily transitioned to more sustainable sources of energy, about 30% of global GHG emissions which come from industrial processes such as cement manufacture, iron and steel production, refineries and petrochemical processing are difficult to decarbonize. CCS offers a promising technique for decarbonizing these hard-to-abate point source emissions. The two key challenges in the implementation of CCS are the availability of safe, secure subsurface storage sites and the high financial costs involved. This study developed low-cost, machine learning-based tools with minimal computational requirements to support the preliminary screening of potential CO_2 storage sites. By training the model on data from natural CO_2 reservoirs, it can predict the suitability of untested sites for CO_2 storage, offering a cost-effective method for identifying promising CCS sites before committing to comprehensive and expensive assessments.

Objectives

This study aims to de-risk CO_2 storage site selection by developing an interpretable, data-driven machine learning framework trained on natural analogues for preliminary assessment of CO2 storage prospectivity of subsurface formations, supporting more reliable CCS deployment.

Method

- Data Collection
  - Data from published literature
- Preprocessing
  - Data cleaning and wrangling
- Dealing with imbalanced data
  - Smote sampling technique
- Model trainning
  - RF
  - XGBOOST
  - MLP
- Cross validation
  - StratifiedKfold validation
- Model selection
  - Precision
  - Recall
  - Accuracy
- Model explanation
  - SHAP values

Results

1

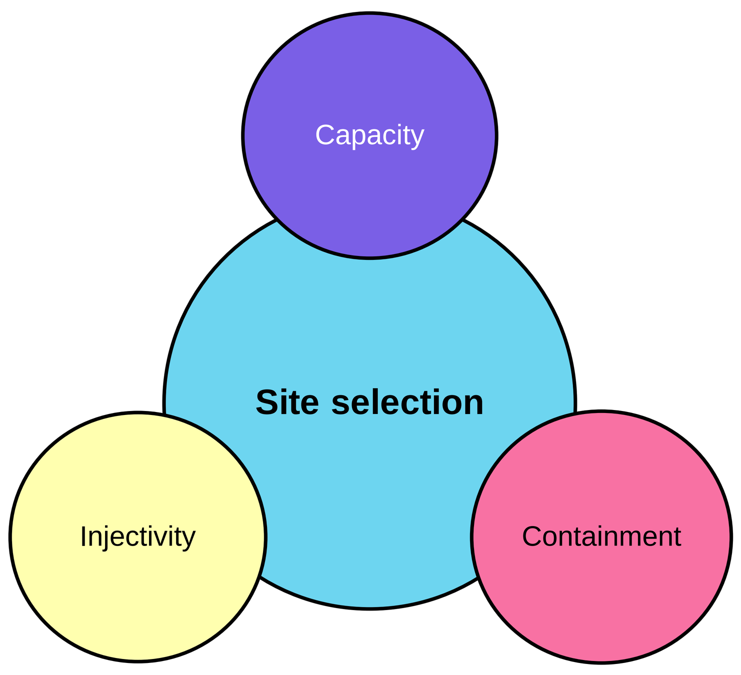

Figure 1. Criteria for Storage Site Selection

4

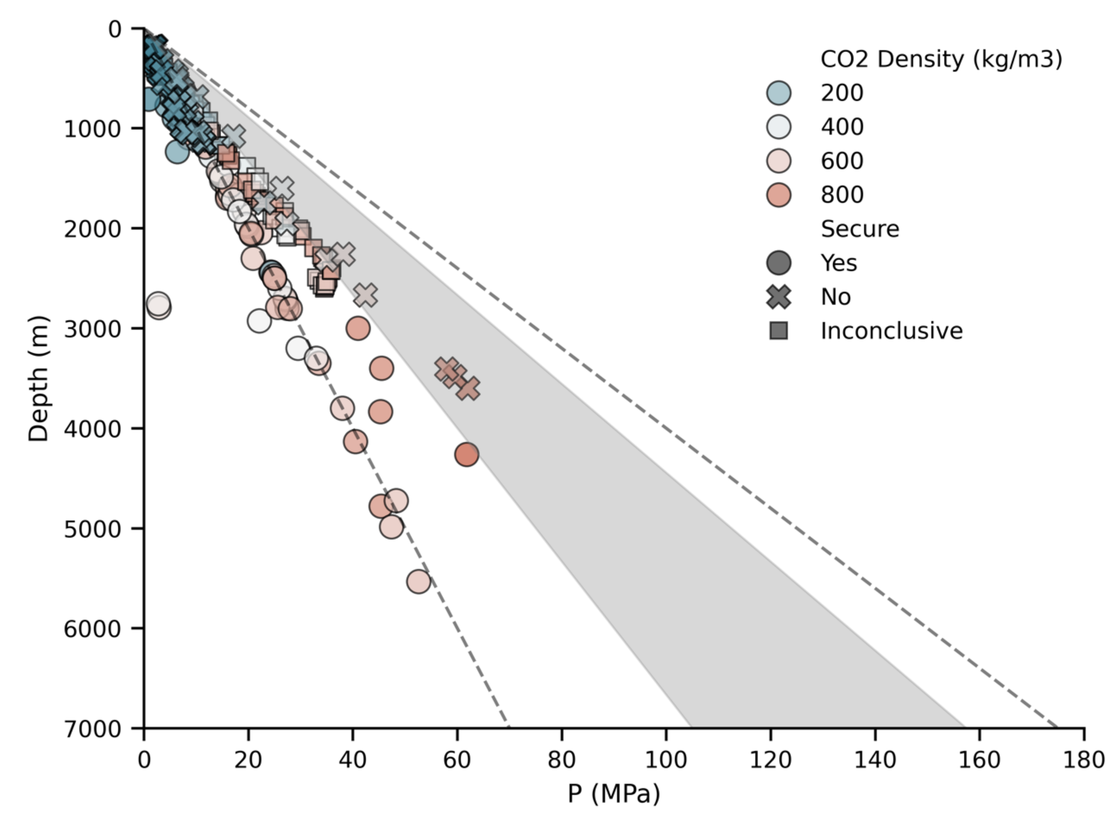

Figure 4. Effect of Pressure, Depth, CO2 Density on storage Security

2

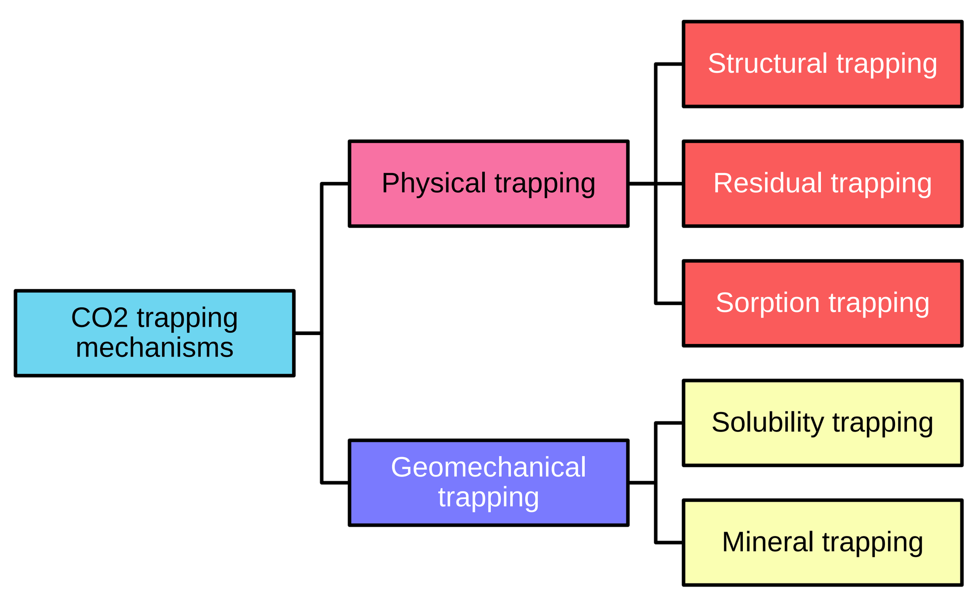

Figure 2. Trapping Mechanisms for Geological CO_2 Storage

5

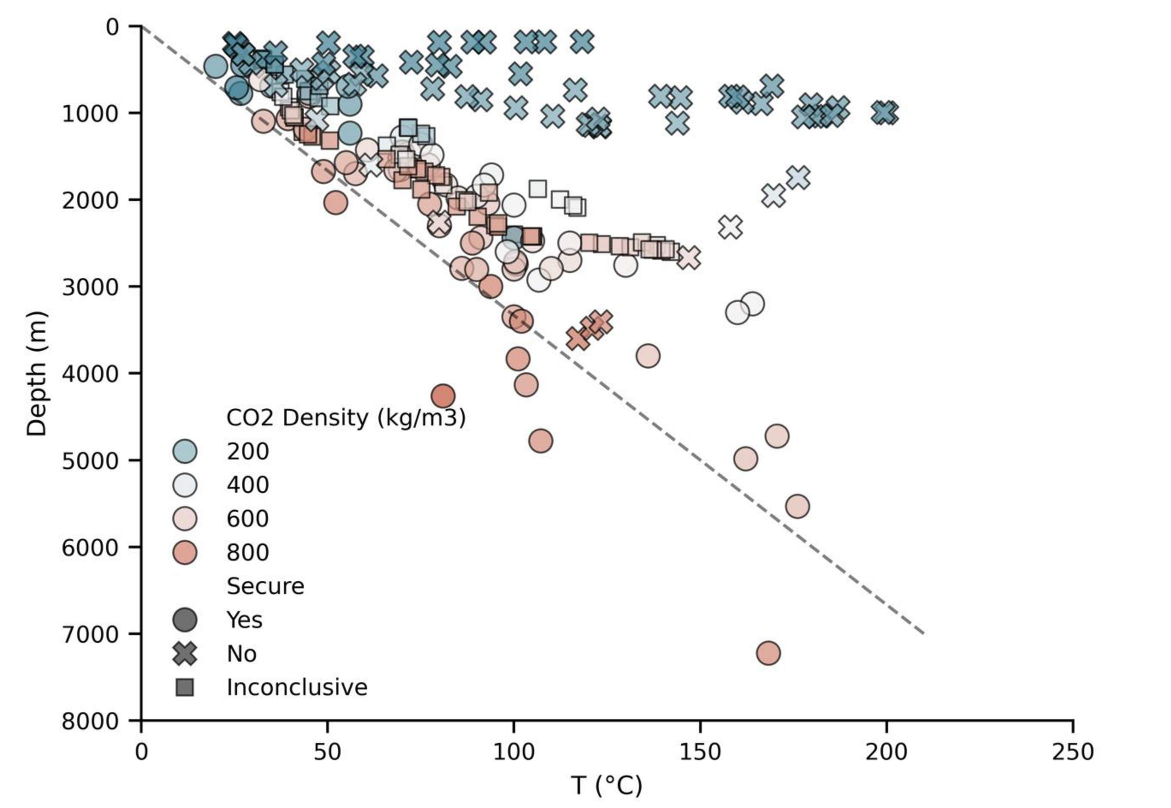

Figure 5. Effect of Temperature, Depth, CO2 Density on Storage Security

3

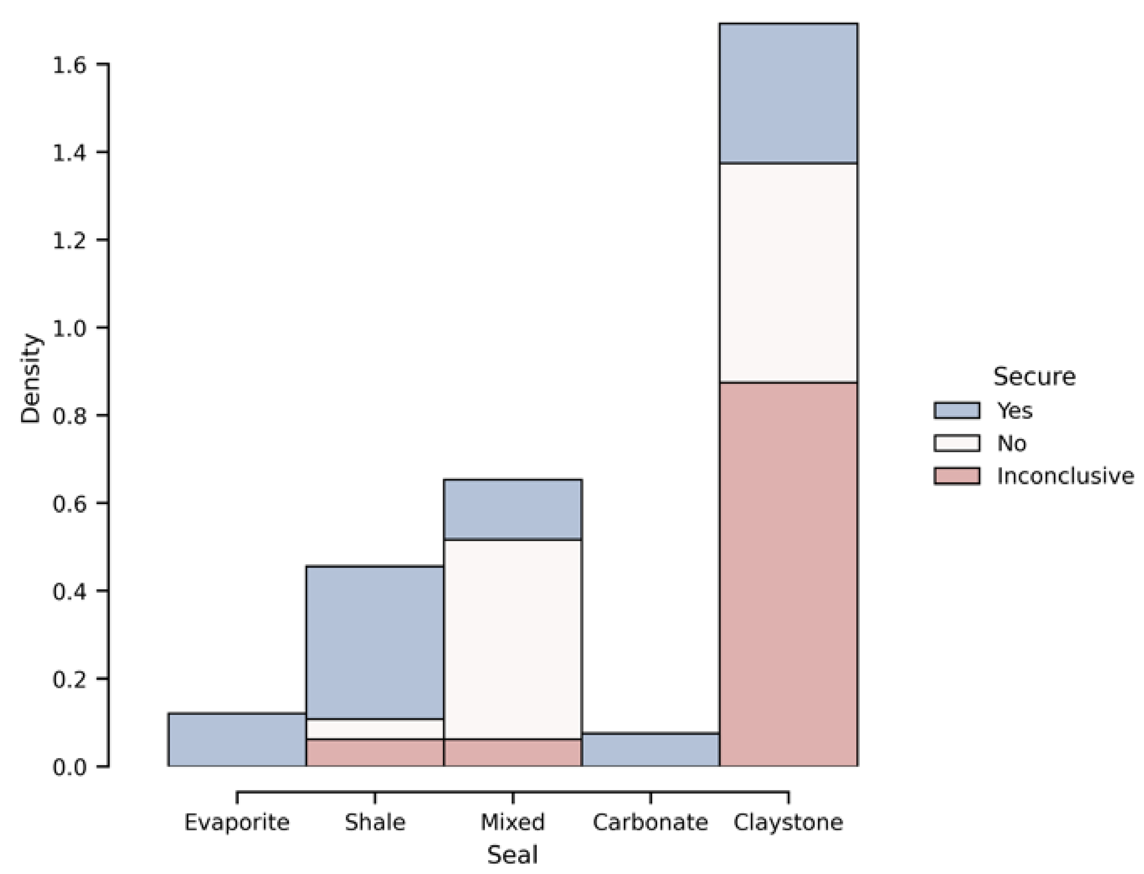

Figure 3. Reservoir Stacking vs. Storage Security

6

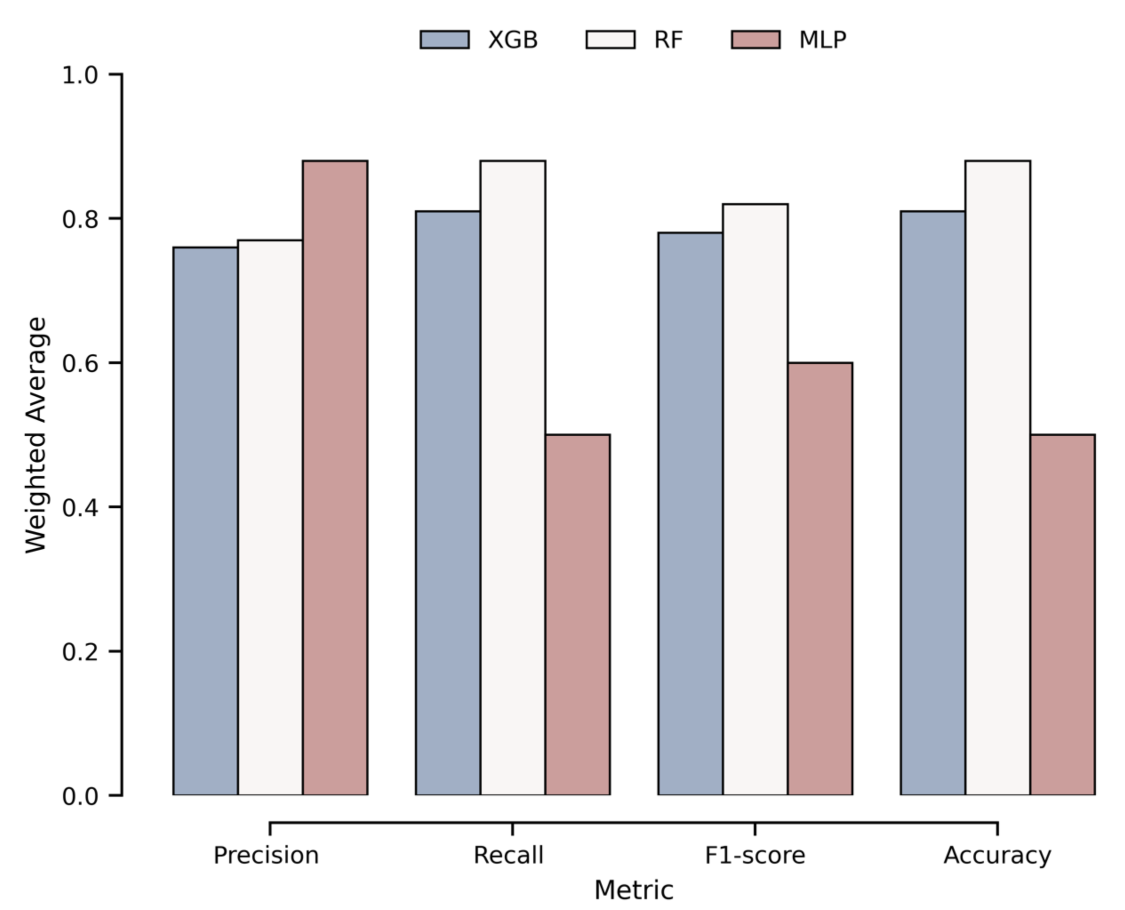

Figure 6. Model Performance on CCS Site Selection

Conclusion

This study developed an interpretable machine learning framework to assess the security of CO_2 storage sites using data from natural analogues. Among the models tested, Random Forest demonstrated the best performance in classification accuracy, recall, and F1-score. SHAP analysis revealed fault presence, seal integrity, and reservoir stacking as key predictors of storage security. The results align with established geological understanding, reinforcing confidence in the model's applicability. This approach offers a scalable, data-driven alternative to subjective assessments, improving site screening for CCS projects.

Acknowledgements

This project was funded by the AGATE Project under Grant Number: 24_25_RSGW_0017. The authors also wish to express their sincere appreciation to the dedicated research assistants at the Net-Zero Carbon Emission Lab (NCEL) for their invaluable support throughout the study.

References

1. "Controls on CO_2 storage security in natural reservoirs and implications for CO_2 storage site selection" (Miocic et al., 2016)
2. "Estimating geological CO_2 storage security to deliver on climate mitigation" (Alcalde et al., 2018)
3. "Prediction of CO_2 storage site integrity with rough set-based machine learning" (Aviso et al., 2019)
4. "Sub-surface geospatial intelligence in carbon capture, utilization and storage: A machine learning approach for offshore storage site selection" (Nassabeh et al., 2024)
5. "A review of CO_2 storage in geological formations emphasizing modeling, monitoring and capacity estimation approaches" (Ajayi et al., 2019)
6. "The noble gas geochemistry of natural CO_2 gas reservoirs from the Colorado Plateau and Rocky Mountain provinces, USA" (Gilfillan et al., 2008)
7. "Criteria and workflow for selecting saline formations for carbon storage" (Callas et al., 2024)