Machine Learning-Based Prediction of CO_2 Storage Capacity and Injectivity

Ajay Kandpal[1], Pradeep Prajapati[2]

[1]Associate Professor, Department of Computer Sciences, Uttarakhand Technical University, Dehradun, UK

[2]Department of Environmental Sciences, Uttarakhand Technical University, Dehradun, UK

ABSTRACT

Accurate prediction of CO_2 storage capacity and injectivity is fundamental to the technical and economic viability of geological carbon storage (GCS), governing site screening, well design, and long-term project planning across depleted hydrocarbon reservoirs, deep saline aquifers, and geothermal-analogue formations. Conventional approaches to capacity and injectivity estimation rely on physics-based reservoir simulation and empirical volumetric methods, both of which are constrained by high computational cost, extensive data requirements, and limited scalability across the large number of geologic realizations needed for robust uncertainty assessment. Machine learning (ML) has emerged as a powerful complementary approach, capable of learning nonlinear relationships between petrophysical, geomechanical, and operational parameters and their resulting storage capacity or injectivity outcomes directly from simulation or field data, often at a fraction of the computational cost of conventional methods. This review synthesizes recent advances in ML-based prediction of CO_2 storage capacity and injectivity, examining artificial neural network, ensemble tree-based, and hybrid architectures applied to storage efficiency estimation, trapping index prediction, and permeability-based injectivity assessment. It draws on case studies from heterogeneous depleted gas reservoirs, geothermal-analogue petrophysical characterization, and geomechanically informed wellbore stability assessment, and further considers methodological parallels with large-scale AI-driven predictive analytics and economic impact frameworks developed for national healthcare systems. The review concludes by identifying persistent challenges in training data scarcity, cross-site generalizability, and integration of geomechanical constraints, and proposes a research agenda for advancing ML-based capacity and injectivity prediction toward routine use in commercial-scale CO_2 storage site screening and design.

Keywords: machine learning, CO_2 storage capacity, injectivity, artificial neural network, deep saline aquifer, petrophysical characterization, carbon capture and storage

1. Introduction

The technical and economic feasibility of any geological CO_2 storage (GCS) project hinges on two closely related questions: how much CO_2 can a given reservoir ultimately store, and at what rate can that CO_2 be safely injected without exceeding pressure or geomechanical constraints. Storage capacity determines the scale of climate mitigation a given site can deliver, while injectivity governs the number and configuration of wells required, the project timeline, and the overall economics of the storage operation. Both quantities depend on a complex interplay of petrophysical properties—porosity, permeability, relative permeability, capillary pressure—geomechanical constraints on maximum sustainable injection pressure, and trapping mechanisms including structural, residual, solubility, and mineral trapping that govern the ultimate fate of injected CO_2 over decadal to centennial timescales.

Traditionally, storage capacity and injectivity have been estimated through a combination of volumetric calculation methods and physics-based reservoir simulation, the latter offering high fidelity but at substantial computational cost, particularly when robust estimates require evaluating capacity and injectivity across the large ensembles of geologic realizations needed to characterize uncertainty in heterogeneous formations. Machine learning has emerged as a increasingly important complementary approach, capable of learning the nonlinear mapping between reservoir properties and storage or injectivity outcomes directly from training data generated by simulation or, where available, field observation, enabling rapid screening and uncertainty quantification across far larger scenario spaces than conventional simulation-in-the-loop approaches can practically support.

This review synthesizes recent advances in ML-based prediction of CO_2 storage capacity and injectivity, integrating insights from heterogeneous depleted gas reservoir case studies, geothermal-analogue petrophysical characterization directly relevant to reservoir quality prediction (Yeboah et al., 2026), and machine-learning-enhanced geomechanical characterization relevant to injectivity constraints (Akpabli et al., 2026), alongside the growing body of ML literature applied directly to storage efficiency, trapping index, and injectivity prediction (Kim et al., 2017; Rezaee & Ekundayo, 2022; Vo Thanh et al., 2020, 2022). It further draws methodological parallels with large-scale AI-driven predictive analytics and economic impact frameworks developed for national healthcare systems (Nguyen, 2026c, 2026d), which offer transferable lessons on deploying predictive AI models to inform resource allocation and cost-effective planning decisions at scale. The objective is to provide a comprehensive synthesis for researchers, engineers, and site developers seeking to advance ML-based capacity and injectivity prediction from research demonstration toward standard practice in commercial CO_2 storage site screening and design.

2. The Storage Capacity and Injectivity Prediction Challenge

2.1 Defining Storage Capacity and Injectivity

CO_2 storage capacity refers to the total mass of CO_2 that a given geologic formation can permanently and safely accommodate, governed by pore volume, trapping efficiency, and pressure constraints on the storage complex. Injectivity refers to the rate at which CO_2 can be delivered into the formation per unit of applied pressure differential, governed primarily by near-wellbore permeability, relative permeability effects as CO_2 displaces resident brine or hydrocarbons, and any near-wellbore formation damage or geomechanical alteration. While related, these two quantities can be decoupled: a formation may possess substantial total storage capacity yet suffer from poor injectivity that necessitates an impractically large number of wells to achieve target injection rates, underscoring the importance of predicting both quantities jointly rather than in isolation.

2.2 Limitations of Conventional Estimation Approaches

Volumetric capacity estimation methods, while computationally inexpensive, rely on simplified assumptions about trapping efficiency that may not capture site-specific heterogeneity, while full-physics reservoir simulation, though more rigorous, becomes computationally prohibitive when applied across the large ensembles of geologic realizations required for uncertainty-quantified capacity and injectivity estimates at heterogeneous sites. This computational burden is particularly acute for injectivity assessment, which requires resolving near-wellbore flow dynamics at a spatial resolution substantially finer than that typically used for field-scale capacity estimation.

2.3 The Machine Learning Opportunity

Machine learning models, once trained on a representative ensemble of simulation or field outcomes, can predict storage capacity and injectivity for new geologic and operational scenarios in a fraction of the time required for a full simulation run, enabling the kind of large-scale, uncertainty-quantified scenario screening that robust site characterization and design demand. This capability is particularly valuable for early-stage site screening, where rapid, approximate capacity and injectivity estimates across many candidate sites or geologic realizations are more valuable than a small number of high-fidelity simulations of a single scenario.

3. Machine Learning Architectures for Storage Capacity Prediction

3.1 Artificial Neural Networks for Storage Efficiency Estimation

Artificial neural networks (ANNs) remain among the most widely applied machine learning architectures for CO_2 storage capacity and efficiency prediction, owing to their capacity to learn complex nonlinear relationships between input reservoir and operational parameters and output storage metrics. Kim et al. (2017) developed an ANN-based model to predict storage efficiency in deep saline aquifers, training the network on 150 representative realizations generated through sensitivity analysis of key aquifer characteristics and validating the resulting model against field-scale data, demonstrating that the ANN approach could predict storage efficiency with high accuracy while avoiding the iterative computational burden of repeated numerical simulation. This foundational approach has since been extended and validated across a range of geologic settings, establishing ANN-based storage efficiency prediction as a mature and widely adopted methodology in the field.

3.2 Ensemble Tree-Based Methods for Trapping Index Prediction

Beyond neural network architectures, ensemble tree-based methods—random forest, extreme gradient boosting (XGBoost), and related algorithms—have demonstrated strong performance in predicting CO_2 trapping indices, which quantify the relative contribution of structural, residual, and solubility trapping mechanisms to overall storage security. Vo Thanh et al. (2022) developed a knowledge-based machine learning framework comparing random forest, XGBoost, and support vector regression models for predicting CO_2 trapping performance in underground saline aquifers, finding that the XGBoost model achieved the highest predictive accuracy when validated against reservoir simulation results through comprehensive blind testing, and proposing the resulting model as a robust screening and process planning tool for uncertainty assessment in commercial-scale carbon storage projects.

3.3 Machine Learning for Enhanced Oil Recovery and Residual Oil Zone Storage

CO_2 storage capacity prediction extends beyond saline aquifers to include CO_2-enhanced oil recovery (CO_2-EOR) operations and storage in residual oil zones, where the coupled objectives of oil recovery and CO_2 retention require predictive models capable of jointly forecasting both outcomes. Vo Thanh et al. (2020) applied an artificial neural network to predict the performance of CO_2-EOR and associated storage capacity in residual oil zones, training the model on numerical simulation results incorporating geological and well-operation uncertainty parameters and achieving high predictive accuracy when validated against real residual oil zone sites in the Permian Basin—a geologic setting closely related to the heterogeneous depleted gas reservoir characteristics documented in Permian Basin storage optimization studies more broadly.

3.4 Relevance of Petrophysical Characterization to Capacity Prediction Fidelity

The accuracy of any machine learning-based capacity prediction model is fundamentally bounded by the quality of the petrophysical property data used to characterize the reservoir. Machine learning-enhanced petrophysical workflows applied to the Three Forks geothermal reservoir in the Williston Basin illustrate how well log, core, and seismic attribute data can be integrated to predict porosity, permeability, and mineralogical composition at high spatial resolution (Yeboah et al., 2026), providing exactly the class of high-fidelity input data that storage capacity prediction models such as those of Kim et al. (2017) and Vo Thanh et al. (2022) require to achieve reliable, site-specific accuracy.

4. Machine Learning Architectures for Injectivity Prediction

4.1 Permeability Prediction as the Foundation of Injectivity Assessment

Because injectivity is governed primarily by near-wellbore permeability, machine learning-based permeability prediction from well log and core data constitutes a foundational capability for injectivity assessment. Rezaee and Ekundayo (2022) applied and compared four machine learning techniques—random forest, artificial neural network, gradient boosting regressor, and support vector regressor—to predict permeability of the Precipice Sandstone in the Surat Basin, Australia, from multiple wireline logs, demonstrating that models trained on a complete set of well logs could generate reservoir permeability estimates with coefficients of determination exceeding 90%, directly informing CO_2 injectivity assessment for this candidate storage formation. This log-to-permeability prediction workflow provides a template directly transferable to other candidate storage sites where extensive well log coverage exists but core-derived permeability measurements remain comparatively sparse.

4.2 Injectivity Assessment in Heterogeneous, Depleted Reservoirs

Depleted hydrocarbon reservoirs present distinct injectivity considerations relative to virgin saline aquifers, given the altered pressure and saturation history left by prior hydrocarbon production. Machine learning-based approaches to injectivity assessment in such settings must account for this production history alongside conventional petrophysical inputs, a challenge directly relevant to heterogeneous depleted gas reservoir storage optimization efforts such as those characterized in the Evetts site case study in the Permian Basin, where substantial spatial variability in porosity, permeability, and residual saturation governs both storage capacity and achievable injection rates across the storage complex.

4.3 Geomechanical Constraints on Injectivity

Injectivity cannot be assessed independently of the geomechanical constraints that bound maximum sustainable injection pressure without risking caprock fracturing, fault reactivation, or wellbore instability. Machine-learning-enhanced geomechanical characterization and wellbore stability assessment, as demonstrated in the offshore South Tano Field, integrates stress field estimation, rock strength prediction, and drilling parameter analysis into a proactive risk-management workflow (Akpabli et al., 2026), providing a geomechanically grounded upper bound on injection pressure that any injectivity prediction model must respect to avoid recommending operationally unsafe injection strategies. Coupling permeability-based injectivity models such as that of Rezaee and Ekundayo (2022) with geomechanically informed pressure constraints represents an important integration point for producing injectivity predictions that are both accurate and operationally safe.

4.4 Joint Prediction of Capacity and Injectivity

Because storage capacity and injectivity are governed by overlapping but distinct sets of reservoir properties, an emerging direction in the literature involves joint, multi-output machine learning models capable of simultaneously predicting both quantities from a shared set of geologic and operational inputs, allowing site screening workflows to directly identify formations offering a favorable combination of high storage capacity and adequate injectivity, rather than evaluating the two metrics through separate, potentially inconsistent modeling pipelines.

5. Integration Across the Site Characterization and Design Workflow

5.1 From Petrophysical Characterization to Capacity and Injectivity Prediction

A coherent ML-based site characterization workflow begins with high-resolution petrophysical characterization—of the kind demonstrated for the Three Forks geothermal reservoir (Yeboah et al., 2026)—which supplies the porosity, permeability, and mineralogical inputs required by both storage efficiency models (Kim et al., 2017) and permeability-based injectivity models (Rezaee & Ekundayo, 2022). This sequential dependency underscores that ML-based capacity and injectivity prediction cannot be treated as a standalone modeling exercise but must be integrated within a broader site characterization pipeline that begins with robust petrophysical data.

5.2 Geomechanical Risk-Constrained Capacity and Injectivity Optimization

Optimal site design requires balancing storage capacity and injectivity objectives against geomechanical risk constraints, a multi-objective optimization problem for which machine learning-enhanced geomechanical characterization (Akpabli et al., 2026) provides essential constraint information. Injection strategies that maximize predicted storage capacity without regard to geomechanical stability risk recommending operationally unsafe designs, underscoring the necessity of embedding geomechanical risk assessment directly within capacity and injectivity optimization workflows rather than treating it as a separate, post-hoc validation step.

5.3 Heterogeneous Reservoir Optimization as an Integrated Case Study

Heterogeneous depleted gas reservoir storage optimization, as exemplified by the Evetts site case study in the Permian Basin, illustrates the practical integration of capacity, injectivity, and geomechanical considerations within a single site-specific optimization framework, where substantial spatial variability in reservoir properties necessitates machine learning-accelerated evaluation of a large number of candidate injection strategies to identify configurations that jointly maximize storage efficiency and injectivity while respecting geomechanical safety margins.

6. Cross-Domain Lessons: Predictive AI and Economic Analysis for Resource Allocation at National Scale

While the technical foundations of ML-based capacity and injectivity prediction are rooted in subsurface petrophysics and reservoir engineering, valuable methodological lessons can be drawn from large-scale predictive AI and economic impact frameworks developed in other data-intensive, resource-allocation-driven domains. National-scale predictive analytics frameworks applied to Medicare and Medicaid populations demonstrate how AI-driven models can identify high-value targets for resource allocation—in that context, high-risk patient populations warranting targeted intervention—from large, heterogeneous datasets characterized by incomplete and variable-quality input information (Nguyen, 2026d). This risk-stratification and resource-prioritization logic maps directly onto CO_2 storage site screening, where ML-based capacity and injectivity predictions serve an analogous function: rapidly identifying, from a large candidate pool of potential storage sites or reservoir zones, those combinations of properties most likely to yield high storage value, thereby directing more expensive, high-fidelity simulation and field characterization resources toward the most promising candidates.

Economic impact analyses of AI adoption in healthcare, quantifying national-scale cost savings and productivity gains attributable to AI-driven predictive modeling, offer a further template for constructing the economic case for ML-based capacity and injectivity prediction investment in CCUS (Nguyen, 2026c). Just as AI-driven healthcare predictive analytics have been shown to reduce costs by enabling more efficient, targeted resource allocation relative to uniform, undifferentiated intervention strategies, ML-based storage capacity and injectivity prediction can be expected to reduce the substantial costs associated with full-physics simulation-based site screening across large candidate portfolios, while simultaneously improving the speed and consistency with which promising storage sites are identified—an economic argument that merits explicit quantification as ML-based prediction tools mature toward standard practice in commercial CO_2 storage site development.

Collectively, these cross-domain parallels reinforce a broader principle: the value of machine learning for rapid, uncertainty-aware prediction in support of resource allocation and screening decisions is not unique to subsurface CO_2 storage, but reflects a general pattern observed across domains characterized by large candidate populations, heterogeneous and costly-to-acquire input data, and the need to prioritize limited downstream investigative or intervention resources. The maturation of ML-based capacity and injectivity prediction for GCS can accordingly draw on economic justification and deployment frameworks already developed in these adjacent, more mature large-scale AI application domains.

7. Challenges and Limitations

Despite substantial progress, several challenges continue to constrain the widespread deployment of ML-based CO_2 storage capacity and injectivity prediction.

Training data scarcity and simulation dependence. Most ML-based capacity and injectivity models are trained on simulation-generated data rather than extensive field observations, given the limited number of commercial-scale CO_2 storage projects with sufficient monitoring history to date, raising questions about how well models trained predominantly on synthetic data generalize to real field conditions.

Cross-site generalizability. Models trained on one geologic setting—such as a specific saline aquifer formation or depleted gas reservoir—may not generalize reliably to structurally or mineralogically distinct settings without substantial retraining, complicating efforts to develop broadly applicable, transferable capacity and injectivity prediction tools across diverse storage portfolios spanning saline aquifers, depleted reservoirs, and geothermal-analogue formations.

Integration of geomechanical constraints. Many existing capacity and injectivity prediction models focus on flow and petrophysical properties in isolation, without explicitly incorporating the geomechanical pressure and stability constraints that ultimately bound safe, achievable injection rates, limiting the direct operational applicability of predictions that have not been validated against geomechanical risk assessment.

Joint versus decoupled prediction of capacity and injectivity. Because storage capacity and injectivity are governed by overlapping but distinct property sets, models that predict one quantity in isolation risk recommending site configurations that are favorable on one metric but deficient on the other, underscoring the need for more widespread adoption of joint, multi-output prediction frameworks.

Model interpretability and site-developer trust. As with other applications of machine learning in high-consequence subsurface engineering, the relatively opaque nature of ensemble tree-based and deep learning models can limit the confidence with which site developers and regulators accept ML-generated capacity and injectivity estimates as a basis for investment and permitting decisions, motivating continued integration of explainability techniques alongside predictive model development.

8. Future Directions

Future research should prioritize the development of joint, multi-output machine learning frameworks capable of simultaneously predicting storage capacity, injectivity, and geomechanical risk from a shared set of petrophysical and operational inputs, building on the individual predictive capabilities demonstrated for storage efficiency (Kim et al., 2017), trapping index (Vo Thanh et al., 2022), residual oil zone storage (Vo Thanh et al., 2020), and permeability-based injectivity (Rezaee & Ekundayo, 2022). Closer integration of high-resolution petrophysical characterization workflows—such as those demonstrated for the Three Forks geothermal reservoir (Yeboah et al., 2026)—directly into capacity and injectivity model training pipelines represents a promising avenue for improving prediction fidelity at data-rich candidate sites.

Extending machine-learning-enhanced geomechanical characterization capabilities—such as those demonstrated for wellbore stability assessment in the offshore South Tano Field (Akpabli et al., 2026)—into unified capacity-injectivity-geomechanics prediction frameworks represents a particularly important direction for ensuring that ML-generated site recommendations remain operationally safe as well as economically favorable, with direct relevance to heterogeneous, multi-constraint storage sites such as the Evetts site in the Permian Basin. Finally, adopting economic justification and resource-allocation frameworks informed by AI-driven predictive analytics deployment in national-scale healthcare systems (Nguyen, 2026c, 2026d) may help establish the cost-benefit case required for ML-based capacity and injectivity prediction to transition from research demonstration toward standard practice in commercial-scale CO_2 storage site screening and design.

9. Conclusion

Machine learning is increasingly central to the prediction of CO_2 storage capacity and injectivity, two quantities that fundamentally govern the technical feasibility and economic viability of geological carbon storage projects. Applications spanning artificial neural network-based storage efficiency estimation, ensemble tree-based trapping index prediction, permeability-based injectivity assessment, and enhanced oil recovery-integrated storage capacity forecasting collectively demonstrate that machine learning can substantially accelerate site screening and design relative to conventional volumetric and full-physics simulation approaches, while enabling the large-scale, uncertainty-quantified scenario evaluation that robust site characterization demands. Cross-domain lessons from AI-driven predictive analytics and economic impact analysis in national-scale healthcare systems further suggest a pathway toward the resource-allocation and cost-justification frameworks necessary for ML-based capacity and injectivity prediction to achieve routine adoption in commercial CO_2 storage practice. Realizing the full potential of machine learning in this domain will require sustained investment in cross-site generalizability, joint capacity-injectivity-geomechanics prediction frameworks, and model interpretability—investments essential to ensuring that geological CO_2 storage sites can be screened, designed, and developed efficiently, safely, and at the scale required to meet global climate objectives.

References

1. Kim, Y., Jang, H., Kim, J., & Lee, J. (2017). Prediction of storage efficiency on CO2 sequestration in deep saline aquifers using artificial neural network. Applied Energy, 185, 916–928. https://doi.org/10.1016/j.apenergy.2016.10.012

2. Nguyen, T. T. (2026c). The economic impact of AI adoption in healthcare: Estimating national cost savings, productivity gains, and long-term health outcomes. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 7(3), 1–11.
3. Nguyen, T. T. (2026d). National-scale predictive analytics for Medicare and Medicaid: An AI-driven approach to identifying high-risk populations and reducing healthcare costs. Journal of Advanced Scientific Research, 17(6), 1–10.
4. Rezaee, R., & Ekundayo, J. (2022). Permeability prediction using machine learning methods for the CO2 injectivity of the Precipice Sandstone in Surat Basin, Australia. Energies, 15(6), 2053. https://doi.org/10.3390/en15062053
5. Vo Thanh, H., Sugai, Y., & Sasaki, K. (2020). Application of artificial neural network for predicting the performance of CO2 enhanced oil recovery and storage in residual oil zones. Scientific Reports, 10(1), 18204. https://doi.org/10.1038/s41598-020-73931-2
6. Vo Thanh, H., Yasin, Q., Al-Mudhafar, W. J., & Lee, K.-K. (2022). Knowledge-based machine learning techniques for accurate prediction of CO2 storage performance in underground saline aquifers. Applied Energy, 314, 118985. https://doi.org/10.1016/j.apenergy.2022.118985