processes

MDPI

Article

Machine Learning for Rapid Evaluation of CO_2 Storage Efficiency in the Weyburn Carbonate Reservoir

Caihua Jia [1,*] and Gang Hui [2,*]

[1] China National Petroleum Corporation Liaohe Oilfield Company Oil Production Technology Research Institute, Panjin 124010, China
[2] State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing 102249, China
* Correspondence: 18624558818@163.com (C.J.); hui.gang@cup.edu.cn (G.H.)

Abstract

Reliable screening of geological CO_2 storage sites requires methods that relate reservoir heterogeneity and injection history to screening indicators without running a full-physics simulation for every candidate scenario. This study presents a machine learning workflow for solubility trapping index (STI)-based screening in the Weyburn Oilfield, Canada, using 77 well records and six geological and operational descriptors: porosity, permeability, rock density, reservoir thickness, injection depth, and injection duration. The residual trapping index (RTI) is retained as a descriptive component of the analytical trapping index framework, while quantitative model comparison focuses on STI. Both indices are treated as well-level analytical screening quantities rather than direct physical measurements. Six regression algorithms were compared under random and spatial block cross-validation. In the displayed random-split comparison, XGBoost achieved the highest R^2 (0.81), but spatial block validation reduced the aggregated R^2 to 0.61, indicating that random-split performance was optimistic due to spatial dependence among nearby wells. Distance-buffer analysis further showed that prediction accuracy declined as training-test separation increased. Feature importance was evaluated using repeated permutation importance with uncertainty intervals and permutation-test *p*-values, supplemented by SHAP analysis to quantify the direction and magnitude of feature contributions. Injection depth and permeability were the most influential predictors in the fitted model, although the ranking is interpreted as an association within the analytical index framework rather than a causal conclusion. The workflow is proposed as a conditional, data-supported screening tool that requires reproducible spatial validation, documented target-index inputs, and further validation against high-fidelity simulation before quantitative storage conclusions are drawn.

Keywords: geological CO_2 storage; carbon capture utilization and storage; Weyburn Oilfield; machine learning; carbonate reservoir; trapping efficiency

Check for updates

Academic Editor: Fabrizio Scala

Received: 20 July 2026
Revised: 30 September 2026
Accepted: 14 October 2026
Published: 2 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Geological storage of CO_2 in deep saline formations and mature hydrocarbon reservoirs is widely regarded as an essential pathway for reducing anthropogenic carbon emissions (Figure 1). The effectiveness of storage, however, depends on coupled processes involving reservoir architecture, multiphase flow, capillary trapping, dissolution, and geochemical interaction. These processes are especially difficult to evaluate in carbonate reservoirs, where pore systems may combine matrix microporosity, vugs, fractures, diagenetic overprints, and internal flow barriers [1,2]. Hydrocarbon reservoirs remain attractive targets for carbon

capture, utilization, and storage (CCUS) because they commonly possess existing geological characterization, proven sealing systems, and surface infrastructure that can shorten the transition from screening to deployment [3]. Field-scale projects have also demonstrated that large volumes of CO_2 can be stored when reservoir and caprock conditions are properly constrained [4].

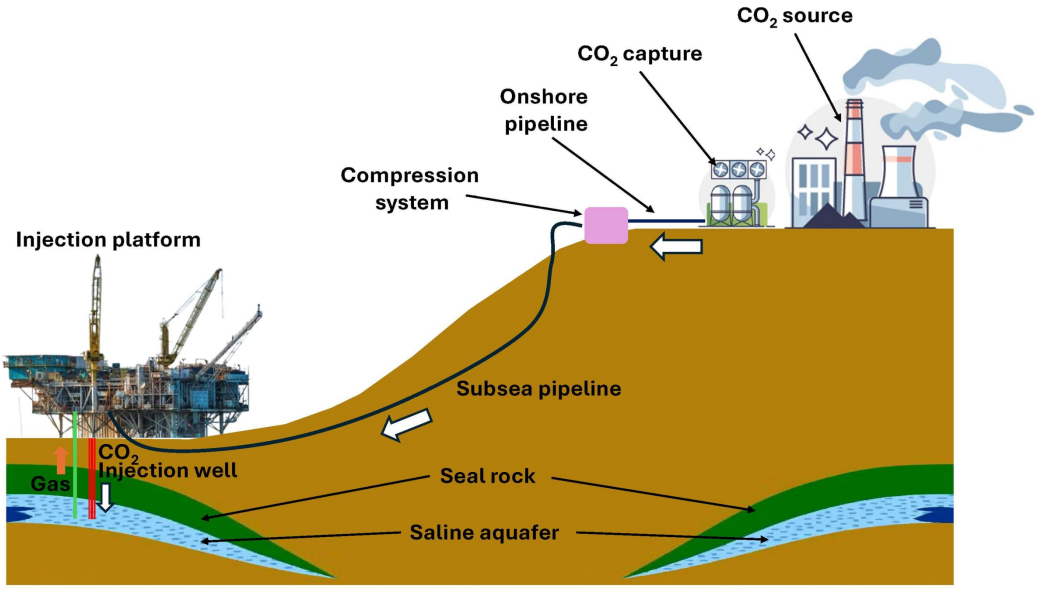


Figure 1. CCUS process chain considered in the geological storage assessment. The schematic links CO_2 capture, transport, utilization, and subsurface storage in depleted oil and gas reservoirs and deep saline aquifers beneath the caprock, together with enhanced oil and gas recovery (EOR/EGR), and highlights the position of storage-efficiency assessment within the broader CCUS workflow. Reprinted/adapted from Ref. [5]; licensed under CC BY 4.0.

A central challenge in storage assessment is to estimate storage efficiency quickly enough for site comparison, scenario screening, and risk-informed project design [6]. High-resolution numerical simulation remains indispensable for detailed project appraisal, but it requires extensive parameterization and repeated computation, which can restrict its use during early-stage screening and sensitivity analysis [7]. The limitation becomes more pronounced in heterogeneous carbonate systems, where trapping behavior is controlled not only by bulk capacity but also by connectivity, stratification, and permeability contrast.

To reduce the computational burden of full-physics simulation, a range of rapid-assessment approaches have been developed. Basin- and reservoir-scale screening and ranking methods use simplified criteria to identify promising storage targets [8,9]. Volumetric and probabilistic screening tools estimate prospective storage resources from static capacity and efficiency factors [10]. Multi-criteria decision analysis has been applied to integrate geological, technical, economic, and risk attributes into site-selection rankings [11,12]. System-level and reduced-complexity models approximate the dominant physical processes with reduced spatial or temporal detail [13]. Response-surface methods replace portions of the simulation chain with analytical or statistical surrogates to accelerate uncertainty propagation and risk assessment [14–16]. These methods are computationally efficient for comparing many candidate sites, but they generally aggregate heterogeneous geological and operational information into fixed criteria, weights, or efficiency ranges, and they do not fully resolve spatial plume evolution, dynamic multiphase interactions, or coupled thermo-hydro-geochemical feedbacks. Consequently, they involve an information trade-off: screening speed is gained at the expense of process fidelity, and the resulting predictions remain conditional on the assumptions and parameter ranges used in their construction.

This trade-off provides a clear logical premise for introducing machine learning as a complementary tool that can learn nonlinear mappings from available field or simulation data while retaining rapid prediction capability.

The long-term stabilization of injected CO_2 is governed by structural, residual, solubility, and mineral trapping [17]. Among these mechanisms, residual and solubility trapping are particularly useful for evaluating the transition from mobile plume migration to more stable storage states. Accordingly, the residual trapping index (RTI) and solubility trapping index (STI) provide compact, process-relevant targets for data-driven prediction [18]. Building reliable predictive models requires datasets that combine geological descriptors, such as porosity, permeability, density, and thickness, with operational variables such as injection depth and injection duration.

Machine learning provides a practical route for complementing physics-based assessment when sufficient field or simulation data are available [19–21]. Data-driven models can approximate nonlinear mappings between multiple input variables and storage-performance indicators at a fraction of the computational cost of repeated numerical simulation [22–24]. When combined with model interpretation, these methods can also reveal which geological and operational attributes exert the strongest influence on predicted storage efficiency, thereby supporting both rapid prediction and process understanding [25–28].

This work uses the Weyburn-Midale CO_2 Monitoring and Storage Project as a field-based case for developing a machine learning screening framework. The novelty of this work is as follows:

(1) Field-specific integration for carbonate CO_2 storage screening. This work assembles and quality-controls a 77-well dataset from the Weyburn carbonate reservoir and uses it to predict the solubility trapping index (STI) within an explicitly disclosed analytical screening framework. This is not a generic tabular dataset, but a field-based dataset with geological and operational descriptors specific to a heterogeneous carbonate system.

(2) Spatial validation rather than only random splitting. We show that the common random 80:20 split produces optimistic performance and that spatial block cross-validation and distance-buffer analysis substantially reduce the reported R^2. To our knowledge, this explicit spatial validation has not been systematically applied in previous machine learning-based screening studies for the Weyburn field.

(3) Transparent treatment of circularity between target and predictors. We openly discuss how STI is computed using an analytical framework that shares some inputs with machine learning predictors, and we therefore interpret the feature ranking as a conditional association within that framework rather than as an independent causal discovery. This is rarely addressed explicitly in similar screening studies.

(4) Interpretability with repeated permutation importance and SHAP. Instead of a single gain-based importance chart, we provide repeated permutation importance with uncertainty intervals and *p*-values, complemented by SHAP and dependence plots. This goes beyond the standard feature-importance output and supports the "interpretable" claim. The contribution is not a new machine learning algorithm or a causal attribution method; it is the integration of data curation, algorithm comparison, and cautious geological interpretation for preliminary CCUS screening.

2. Study Area and Data Context

The Weyburn Oilfield is situated in the Williston Basin of southeastern Saskatchewan, Canada, and is one of the best-documented sites associated with the Weyburn-Midale CO_2 Monitoring and Storage Project (Figure 2). Production occurs mainly from the Middle Beds of the Mississippian Charles Formation, a carbonate–evaporite succession deposited in a

restricted shallow-marine platform setting [18,29]. The reservoir interval is overlain by the regionally persistent Middle Evaporite, which forms a competent seal and limits upward fluid migration.

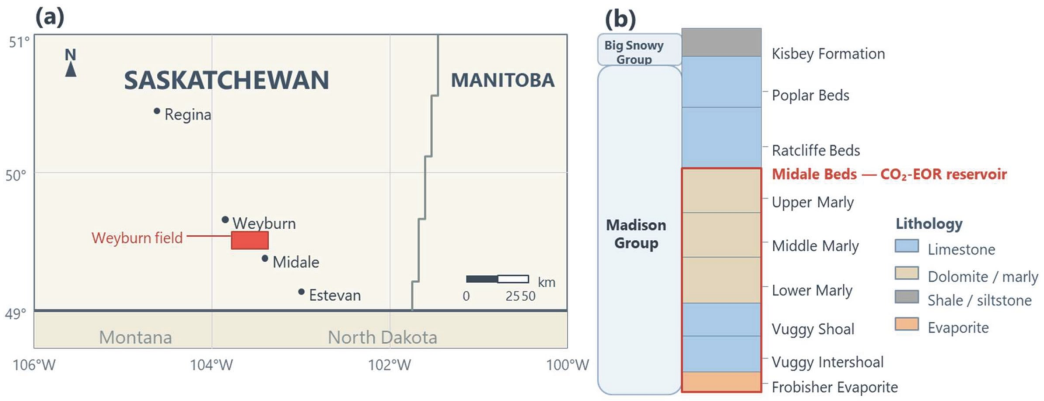


Figure 2. Location and geological framework of the Weyburn study area. (**a**) Position of the Weyburn-Midale field in southeastern Saskatchewan, Canada. (**b**) Stratigraphic context of the Marly and Vuggy reservoir units and the overlying evaporite sealing interval. Original artwork.

The storage interval is commonly divided into the Marly and Vuggy units, which differ substantially in pore structure and flow capacity. The Marly unit is dominated by microcrystalline dolostone with relatively high porosity but low permeability, reflecting deposition and diagenesis in a low-energy evaporitic setting. The Vuggy unit contains skeletal limestone with dissolution-enhanced vuggy porosity and locally connected high-permeability pathways. This contrast produces a vertically heterogeneous system in which CO_2 migration, brine contact, and trapping are controlled by both pore volume and connectivity [18].

For the Weyburn area, representative regional background values are as follows: the geothermal gradient is approximately 25–30 °C/km, the hydrostatic pressure gradient is approximately 9.8–10.5 MPa/km, and the surface temperature reference is approximately 4–10 °C. These regional approximations are consistent with the measured depth range of 704–3853 m, which is used only to describe the well paths, not the reservoir depth. The reservoir TVD range is approximately 1200–1800 m, with a mean of about 1476 m, which is consistent with the reported pressure range. The reported pressure range is 12.5–18 MPa. Well-specific pressure and temperature data should be used when available [18,29,30].

Reservoir quality in Weyburn reflects a complex diagenetic history. Dolomitization preserved and enhanced microporosity in the Marly interval, whereas cementation, dissolution, and localized fracturing contributed to the more heterogeneous pore network of the Vuggy interval [30]. Core observations indicate that permeability may be concentrated in thin flow units or connected vug–fracture networks. Such heterogeneity is directly relevant to CO_2 storage because it influences sweep, plume geometry, contact between CO_2 and brine, and the distribution of residual trapping.

The dataset used in this study integrates information from 77 wells with sufficiently complete records of reservoir properties and CO_2 injection history. The selected variables include porosity, permeability, thickness, rock density, injection depth (i.e., reservoir TVD), and injection duration. Well information was obtained from the publicly available PetroNinja database (https://petroninja.com/wells, accessed on 1 October 2026), which compiles well logs, core measurements, and operational records from Canadian energy datasets. The Weyburn field has been investigated in previous peer-reviewed studies, and a subset of the compiled records was cross-checked against published information to improve internal consistency [29,30]. The porosity range of 0.03–0.20 in Table 2 is the compiled dataset range derived from core and log interpretations, whereas the Marly 16–38% and Vuggy 8–20%

ranges in Table 2 are regional literature ranges; the two are therefore not identical, and the compiled dataset falls largely within the Vuggy range.

The selected wells cover a measured depth range of 704–3853 m and capture broad lateral and vertical variation in reservoir properties. Figure 3 summarizes the spatial distribution of the wells, whereas Table 3 reports the statistical range of the input and output variables. The resulting dataset provides a compact but geologically meaningful basis for examining how storage-efficiency indicators respond to reservoir heterogeneity and injection conditions. The predictor thickness is the total reservoir thickness of the completed interval reported for each well, whereas the interval of 18–29 m in Table 2 refers to the literature stratigraphic thickness of the Marly and Vuggy units; the compiled dataset includes intervals that locally exceed the type-section thickness, as shown by the range of 0.5–79.1 m in Table 3.

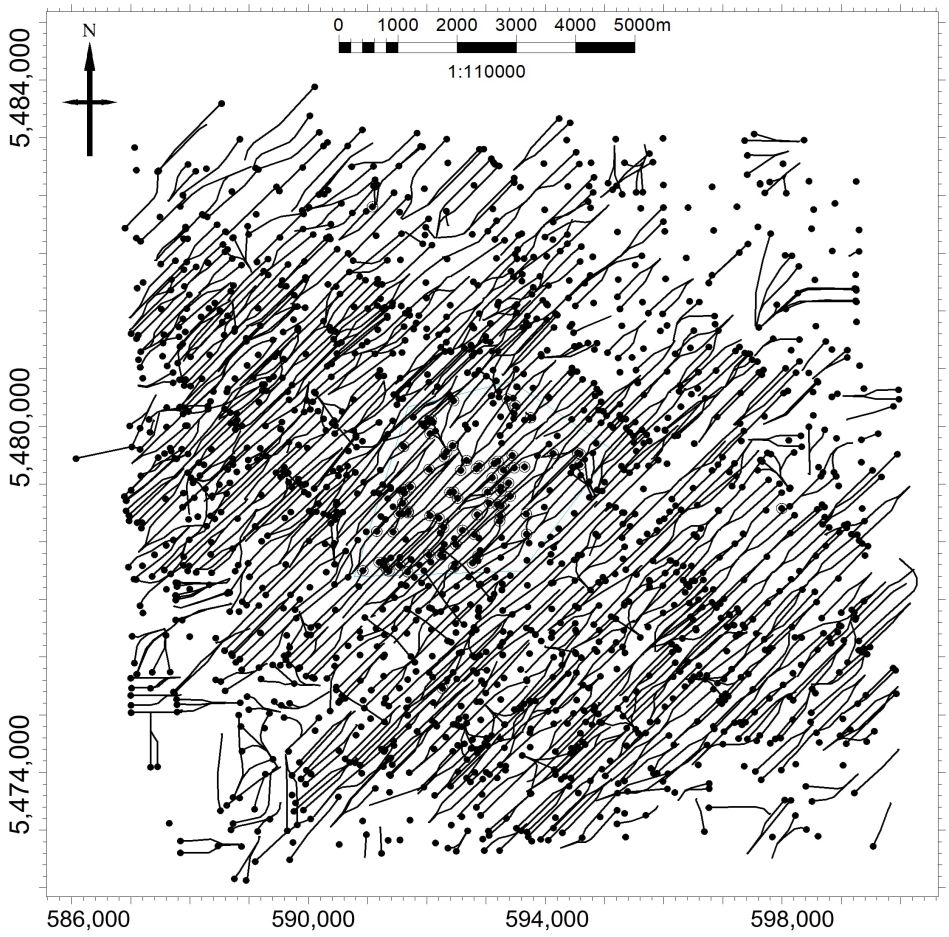


Figure 3. Spatial distribution and well geometry in the Weyburn dataset. Black dot denotes the wellhead of the well. The black line represents the well path of the horizontal well.

3. Materials and Methods

The proposed workflow couples field-data integration, preprocessing, machine learning prediction, and geological interpretation (Figures 4 and 5). First, reservoir and injection variables were compiled from multiple well-level sources and screened for completeness. Second, the dataset was cleaned to remove inconsistent entries and to harmonize units. Third, the data were divided into training and independent test subsets, with 80% of the samples used for model development and 20% withheld for final performance evaluation [31]. This design was selected to provide a reproducible comparison among algorithms while maintaining a separate dataset for unbiased testing.

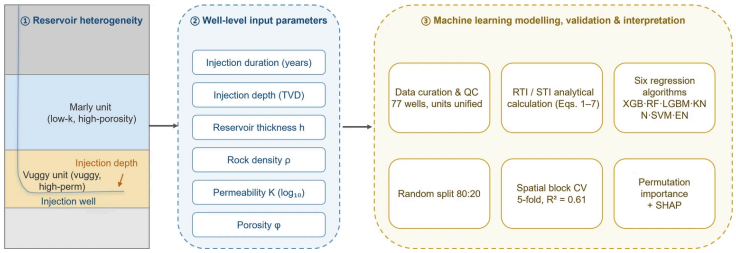


Figure 4. Conceptual framework for machine learning-based CO_2 storage-efficiency screening in a carbonate reservoir. The workflow integrates reservoir-scale descriptors and CO_2 injection-volume information.

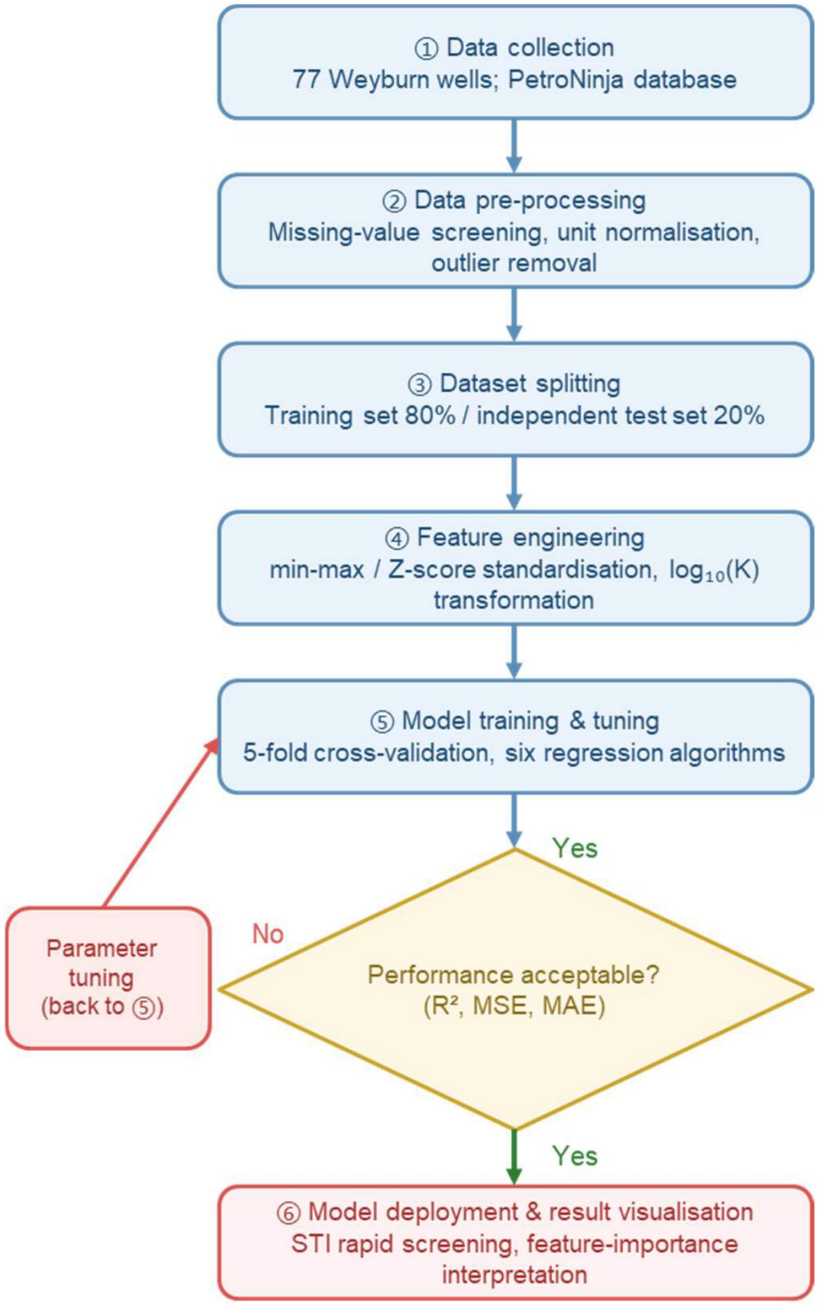


Figure 5. Workflow for model development and application. The procedure comprises data collection, preprocessing, train–test splitting, model training, and compliance assessment.

To reduce overfitting and improve model robustness, cross-validation was applied during model tuning [25,26]. Six algorithms representing ensemble, kernel, similarity-based, and regularized-linear approaches were trained and compared [5]. Predictive skill was quantified using the coefficient of determination (R^2) and mean squared error (MSE) [31]. The model that combined the highest R^2 with the lowest MSE on the independent test set was then interpreted to identify the geological and operational controls on CO_2 storage efficiency.

3.1. Dataset Construction and Preprocessing

The database was assembled from public Weyburn well information, including core-derived reservoir properties, digital well-log interpretations, and available CO_2 injection records (Figure 6, Table S1). The PetroNinja well database was used as the primary source for well-level information, and the compiled records were harmonized to generate a consistent set of predictors [29]. The reservoir TVD range is approximately 1200–1800 m, with a mean of about 1476 m, corresponding to intervals that capture the principal reservoir variability relevant to CO_2 storage evaluation. The injection depth (i.e., reservoir TVD) is regarded as an operational variable that serves as a proxy for pressure–temperature conditions.

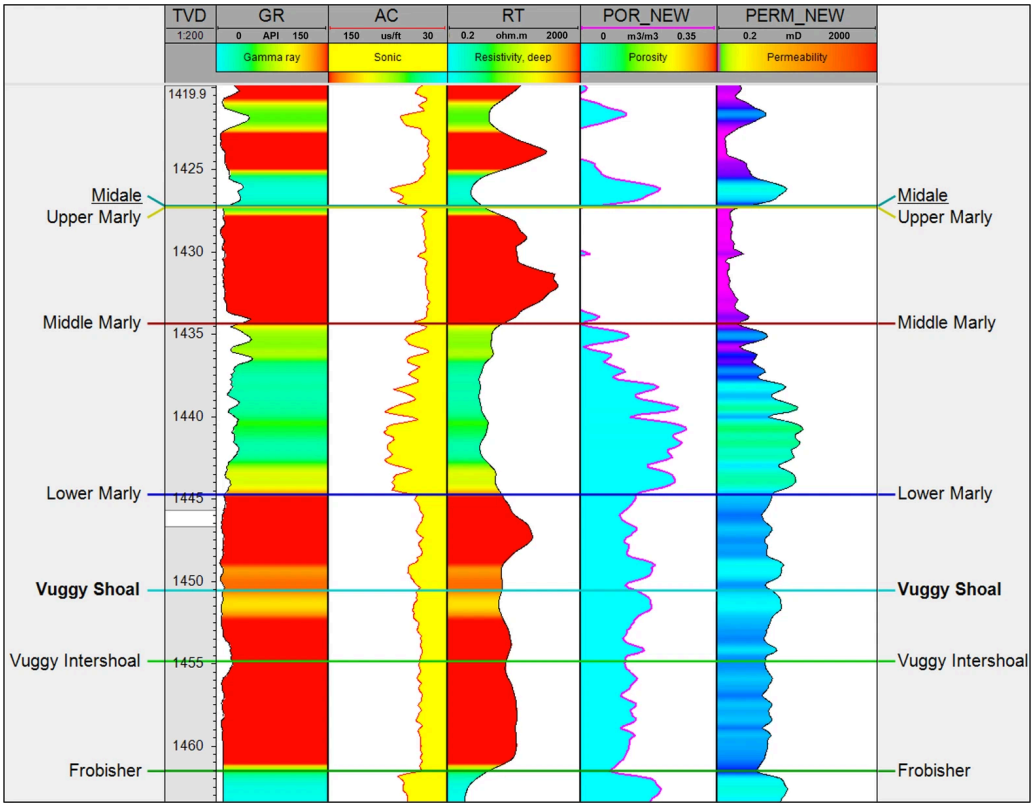


Figure 6. Reservoir-property profiles for the representative well used in data compilation. GR, gamma ray; AC, sonic; RT, deep resistivity; POR_NEW, porosity; Perm_NEW, permeability.

A strict inclusion rule was applied to ensure that each selected well contained both geological descriptors and injection information required for target-index calculation. Wells were retained only when records were available for all six input features, when the injection data were sufficient to calculate RTI and STI, and when the core or log interpretations were internally consistent. After this screening step, 77 wells remained in the final dataset.

The predictor set was designed to represent the main static and operational controls on storage efficiency: reservoir thickness, rock density, porosity, permeability, CO_2 injection depth, and cumulative injection duration. Injection duration is treated as a cumulative vari-

able measured up to the time represented by each well record. This choice is appropriate for the present cross-sectional analysis because the predicted trapping indices summarize time-integrated storage behavior. Nevertheless, the injection process is inherently dynamic; future time-resolved studies could therefore use recurrent or transformer-based architectures to model the temporal evolution of plume migration and trapping.

Several potentially important variables were not included because they were unavailable or inconsistently recorded across the selected wells. These include formation-water salinity, reservoir temperature, and time-varying injection rate. Their absence should be considered when interpreting the model because these parameters can influence CO_2 density, dissolution, capillary behavior, and geochemical reactions. Incorporating these variables in future datasets is expected to improve both prediction accuracy and physical interpretability.

For each well i, the cumulative net injected CO_2 mass is defined as:

$$M_{inj,i} = \sum_t \left(\dot{m}_{co_2,i,t} \Delta t - M_{rec,i,t} \right), \quad (1)$$

where $\dot{m}_{co_2,i,t}$ is the metered CO_2 injection mass rate, and $M_{rec,i,t}$ is the recovered or recycled CO_2 mass, included only when the reported injection ledger is not already net of recycling. When injection is reported as standard volume, mass is converted using the documented standard-state temperature and pressure, CO_2 purity, and the corresponding standard-state CO_2 density:

$$M_{inj,i} = \sum_t V_{std,i,t} \rho_{CO_2}, stdx_{CO_2,i,t}, \quad (2)$$

The response variables were the solubility trapping index (STI) and residual trapping index (RTI), defined in Equation (3). In this study, they are analytical screening indices rather than direct field measurements or outputs from a full-physics simulator. The model therefore approximates the adopted index framework and must not be interpreted as an independent measurement of physical dissolved or residually trapped CO_2 mass. The solubility trapping index and residual trapping index are then calculated as:

$$STI_i = \frac{M_{diss,i}}{M_{inj,i}}, \; RTI_i = \frac{M_{res,i}}{M_{inj,i}}, \quad (3)$$

The dissolved CO_2 mass is calculated from a pressure-, temperature-, and salinity-dependent CO_2–brine solubility correlation:

$$M_{diss,i} = C_{CO2}(P_i, T_i, S_i) V_{brine,contact,i}, \quad (4)$$

$$V_{brine,contact,i} = f_{contact,i} V_{pore,i} S_{w,i}, \quad (5)$$

$$V_{pore,i} = V_{bulk,i} \emptyset_i \quad (6)$$

Here, C_{CO2} is evaluated using the stated CO_2–brine solubility correlation, V_{bulk} is the representative well drainage or grid volume, $\emptyset$ is porosity, Sw is water saturation, and $f_{contact}$ represents the fraction of formation water contacted by CO_2. A correlation applicable to the reported pressure–temperature–salinity range must be identified explicitly; for example, the Duan and Sun model covers CO_2 solubility in pure water and aqueous NaCl solutions over broad pressure, temperature, and salinity ranges. The residual CO_2 mass is calculated as:

$$M_{res,i} = \rho_{CO2}(P_i, T_i) V_{pore,i} f_{swept,i} S_{gr,i}, \quad (7)$$

where ρ_{CO2} is the density of the free-phase CO_2, $f_{swept,i}$ is the swept-pore-volume fraction, and S_{gr} is the residual gas saturation after imbibition. It is clarified that S_{gr} must be taken from Weyburn-specific core-flood/hysteresis measurements or a documented history-matched reservoir model. A generic literature value or a sandstone-derived endpoint is not treated as a Weyburn measurement.

For every input quantity, a new parameter-provenance table reports its value or range, unit, temporal basis, data source, conversion procedure, and uncertainty range (Table 3). The geostatistical model, where used, is limited to spatial aggregation/interpolation of static reservoir properties and representative volumes; it is not a compositional-flow or reactive-transport model and does not itself generate dissolved or residually trapped CO_2 masses.

The present revision does not report a numerical uncertainty propagation because the well-level inputs needed to quantify injection-meter uncertainty, CO_2 purity, pressure, temperature, salinity, contacted volume, residual gas saturation, and in situ CO_2 density are not fully traceable in the submitted dataset. These quantities must be reported, with sources and uncertainty ranges, before the STI or RTI can be used for quantitative storage certification.

Before model training, the input variables were scaled to avoid biases arising from differences in units and numerical ranges. Two common preprocessing options were compared: min–max normalization, which rescales variables to a fixed $[0, 1]$ interval, as shown in Equation (8), and Z-score standardization, which transforms variables according to their mean and standard deviation, as shown in Equations (9)–(11). Summary statistics, including minimum, maximum, mean, standard deviation, skewness, and kurtosis, were calculated to describe the distributions of all variables.

The strongly skewed distributions of permeability, thickness, and injection duration require cautious preprocessing. Linear scaling reduces unit-related numerical imbalance but does not remove the leverage effect of extreme values. A log10 transformation of permeability was therefore applied and evaluated as a robustness test. The comparison of the log-transformed and untransformed permeability predictors, including the stability of model performance and feature ranking, is reported in Table 4 and Section 3.1.

$$y_i = \frac{x_i - \min\limits_{1\le i\le n}(x_i)}{\max\limits_{1\le i\le n}(x_i) - \min\limits_{1\le i\le n}(x_i)} \tag{8}$$

$$\mu = \frac{\sum_{i=1}^{n}(x_i)}{n} \tag{9}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}} \tag{10}$$

$$y_i = \frac{x_i - \mu}{\sigma} \tag{11}$$

In these equations, n denotes the number of samples, x_i is the i-th original input value, and y_i is the corresponding normalized value. The min–max transformation rescales each feature to a fixed interval, whereas Z-score standardization uses the dataset mean, μ, and standard deviation, σ, to center and scale the data.

A logarithmic permeability transformation before scaling was introduced:

$$k_i^{log} = log_{10}\left(\frac{k_i}{1mD}\right), \tag{12}$$

where k_i is the original permeability of well i. The transformed variable was subsequently standardized using statistics calculated from the training subset only:

$$z_i^{log} = \frac{k_i^{log} - \mu_{train}^{log}}{\sigma_{train}^{log}}, \tag{13}$$

This procedure avoids information leakage from the test subset and preserves comparability with the original-permeability workflow. The six models were retrained using the same repeated train–test splits, feature set, tuning procedure, and evaluation metrics. The

R^2, MSE, and MAE between the original-permeability and log-transformed-permeability models are compared.

The results show that the log transformation changed the distributional representation of permeability and reduced its skewness from 2.84 to 0.33. For the XGBoost model, the mean test-set R^2, MSE, and MAE were 0.810, 0.021, and 0.1047 using the original permeability and 0.830, 0.019, and 0.098 using $\log_{10}(K)$ respectively. The corresponding results for the other five models are summarized in Table 4.

Feature importance was further recalculated using repeated permutation importance on the held-out test data, and the resulting rankings before and after transformation were compared. The rank agreement between the two feature representations was quantified using Spearman's rank correlation coefficient. The importance of permeability was stable, whereas the interpretation of all features is now reported with the appropriate caution.

It is noted that tree-based algorithms are expected to be less sensitive to a monotonic transformation of permeability than distance-based, kernel-based, and linear models because tree splits primarily depend on rank order. Nevertheless, the comparison was performed empirically rather than assumed. The revised manuscript therefore reports the log-transformed analysis as a robustness test and bases the final interpretation on the representation that provides the most stable repeated-validation performance.

To represent spatial heterogeneity, a three-dimensional geostatistical model of key reservoir properties was constructed using sequential Gaussian simulation. The simulation honored spatial-correlation structures and porosity–permeability relationships inferred from the multi-well dataset (Figure 7). It is worth noting that the sequential Gaussian simulation model is used only for spatial characterization and to provide the horizontal variogram range used in the distance-buffer sensitivity analysis; it is not used to generate predictor or target values for the machine learning models.

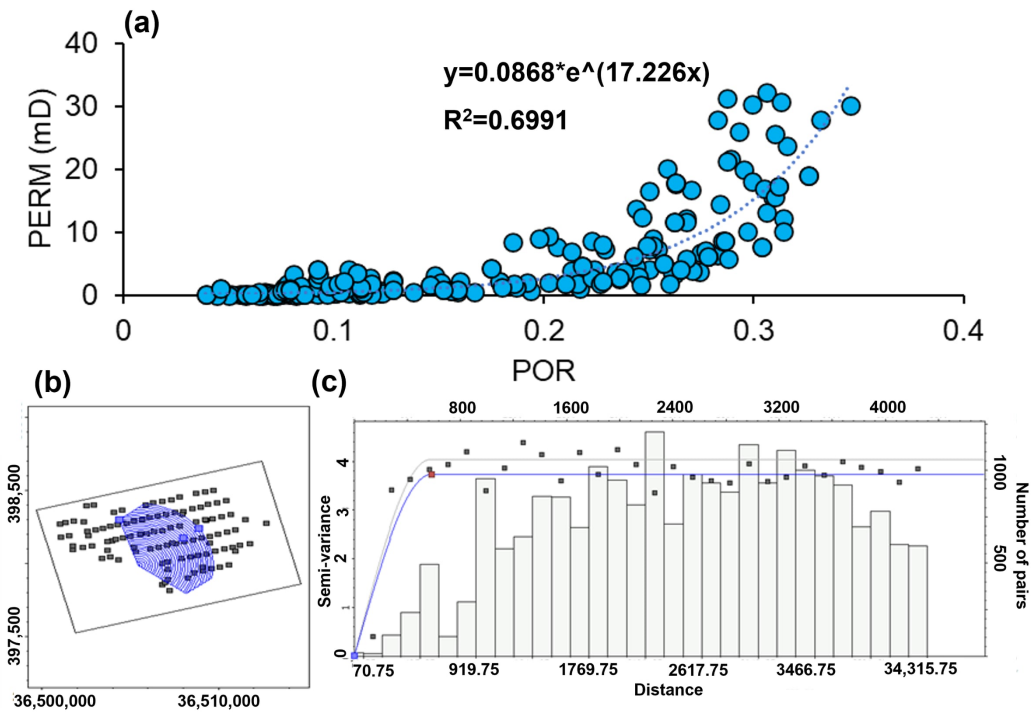


Figure 7. Three-dimensional geostatistical representation of reservoir heterogeneity. (**a**) Core porosity–permeability relationship by lithofacies. (**b**) Variogram analysis in the principal spatial direction is used to constrain sequential Gaussian simulation. (**c**) Experimental semi-variogram of permeability showing the empirical semivariance.

Additionally, Pearson and Spearman correlation matrices were computed for all input variables to examine linear and monotonic associations. Variance inflation factors (VIFs) were also calculated to assess potential multicollinearity, particularly for injection depth, which may serve as a proxy for pressure–temperature conditions.

3.2. Machine Learning Models

The dataset was randomly partitioned into a training subset (80%) and an independent test subset (20%). Six regression algorithms were selected to compare different model structures and inductive assumptions:

Ensemble tree models: XGBoost, Random Forest, and LightGBM. These methods are well suited to nonlinear relationships and interactions among geological and operational variables. They also provide feature-importance measures, which are useful for connecting prediction results with reservoir interpretation.

Classical benchmark models: K-Nearest Neighbors (KNN), Support Vector Machine (SVM) with a radial basis function kernel, and ElasticNet-regularized linear regression. These models provide complementary baselines to assess whether the problem is dominated by local similarity, kernel-transformed relationships, or approximately linear trends.

Table 1 lists the hyperparameters and modeling assumptions for the six machine learning algorithms. Hyperparameters were tuned using cross-validation within the training subset, and the final comparison was performed on the independent test subset.

Table 1. Hyperparameters and modeling assumptions for the six machine learning algorithms.

Model	Type	Principal Hyperparameters Used in This Study	Main Assumptions/Inductive Bias Relevant to the Small-Sample Geological Dataset
XGBoost	Gradient-boosted trees	100 estimators; max depth = 6; learning rate = 0.1; reg_λ = 1.0; reg_α = 0.1	Captures nonlinear relationships and feature interactions. Tree splits depend mainly on rank order, so the model is relatively insensitive to monotonic transformations. Regularization and depth limits are used to reduce overfitting on the compact 77-well dataset.
Random Forest	Bagged decision trees	100 trees; max depth = 10; min samples split = 5	Reduces variance by averaging many decorrelated trees. Robust to noise and less prone to overfitting than a single tree. Assumes that ensemble averaging of tree-based partitions can approximate complex geological relationships.
LightGBM	Histogram-based gradient-boosted trees	31 leaves; learning rate = 0.05; 100 estimators	Efficient gradient boosting with leaf-wise growth. Captures nonlinearity and interactions. The relatively small number of leaves and low learning rate are intended to limit model complexity for a small dataset.

Table 1. *Cont.*

Model	Type	Principal Hyperparameters Used in This Study	Main Assumptions/Inductive Bias Relevant to the Small-Sample Geological Dataset
KNN	Distance-based local learner	5 neighbors; distance-based weighting	Predictions are based on local similarity in feature space. Assumes that wells with similar predictor values have similar STI. Sensitive to feature scaling and irrelevant variables, so standardization is required.
SVM	Kernel-based regression	RBF kernel; C = 1.0; gamma = 'scale'	Maps inputs into a high-dimensional feature space and fits a regularized regression function. Assumes that nonlinear relationships can be captured through the kernel. Performance depends on C and gamma, and on feature scaling.
ElasticNet	Regularized linear regression	L1/L2 mixing ratio = 0.5; alpha = 0.1	Assumes an approximately linear relationship between predictors and STI, with regularization to handle correlated predictors and prevent overfitting. Provides a transparent linear benchmark against the nonlinear models.

3.3. Model Configuration and Evaluation Metrics

For reproducibility, the principal hyperparameters are reported here. XGBoost used 100 estimators, a maximum tree depth of 6, a learning rate of 0.1, reg_lambda = 1.0, and reg_alpha = 0.1. Random Forest used 100 trees, a maximum depth of 10, and a minimum of 5 samples for node splitting. LightGBM used 31 leaves, a learning rate of 0.05, and 100 estimators. KNN used 5 neighbors with distance-based weighting. SVM used a radial basis function kernel with C = 1.0 and gamma = "scale". ElasticNet used an L1/L2 mixing ratio of 0.5 and alpha = 0.1. These initial settings were selected to provide conservative baseline models for a compact dataset: tree depth, regularization, and split constraints were used to reduce overfitting in ensemble models, while KNN, SVM, and ElasticNet were retained as transparent benchmark learners with different inductive assumptions. All models were run with a fixed random seed of 42 and implemented in Python 3.9 using scikit-learn v1.1.2, XGBoost v1.7.0, and LightGBM v3.3.2.

Predictive performance was evaluated using R^2, MSE, and MAE. MSE measures the average squared prediction error and is sensitive to large residuals, MAE reports the mean absolute prediction error on the trapping index scale, and R^2 describes the proportion of target-variable variance explained by the model. The preferred model was therefore selected from the combined evidence of high R^2, low MSE, and low MAE on the independent test data.

Equation (14) defines MSE, which quantifies the mean squared difference between observed and predicted values.

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{n} \quad (14)$$

where y_i and $\widehat{y_i}$ denote the observed and predicted STI values for sample i, respectively, and n is the number of test samples. Lower values indicate better predictive accuracy.

MAE quantifies the mean absolute difference between observed and predicted STI values, as defined by:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \widehat{y_i}|}{n} \quad (15)$$

A lower MSE indicates a smaller prediction error, with zero representing an exact prediction. Equation (16) defines R^2, which measures the fraction of observed variance reproduced by the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \quad (16)$$

where $\overline{y_i}$ is the mean of the observed values. An R^2 value of 1 indicates perfect agreement between prediction and observation, whereas an R^2 value near 0 indicates little improvement over predicting the sample mean.

Because the dataset contains 77 wells and 6 input features, overfitting was addressed through 5-fold cross-validation during hyperparameter tuning. Additional regularization was included in the ensemble models, including reg_lambda, reg_alpha, limited tree depth, and minimum split-size constraints. The final reported performance is based on the 20% hold-out test subset, which was not used during training or cross-validation.

To evaluate the robustness of the models under the limited sample size, a repeated random-splitting validation strategy was further adopted. The full dataset containing 77 wells and six input features was randomly divided into training and test subsets 100 times using independent random seeds. In each repetition, model training, feature standardization when required, and hyperparameter tuning were conducted only on the training subset, whereas the test subset was used exclusively for external performance evaluation. This procedure was designed to reduce the dependence of the reported performance on a single data partition and to assess the stability of the prediction results.

For each random split, the coefficient of determination R^2, mean squared error (MSE), and mean absolute error (MAE) were calculated. The final robustness assessment was reported as the mean ± standard deviation of the 100 test-set results, and the empirical 95% confidence interval was obtained from the 2.5th and 97.5th percentiles of the metric distributions.

4. Results

4.1. Distribution of Reservoir and Operational Variables

Table 2 lists the intermediate physical parameters required for calculating the STI and RTI. Table 3 reports their descriptive statistics. The variables include the RTI and STI as target indicators; geological predictors such as porosity, permeability, thickness, and density; and operational predictors such as injection depth and injection duration.

Figure 8 summarizes the map-view distribution of well-averaged input and output variables. Each symbol represents a well, and the color scale indicates the well-averaged value of the variable rather than a continuous spatial interpolation.

Table 2. Intermediate physical parameters required for calculating STI and RTI.

Parameter	Value/Range	Unit	Uncertainty Range	Data Source
Pressure (P)	12.5–18 MPa, mean ≈ 15 MPa; initial ≈ 14 MPa	MPa	Individual measurement error small (±0.1–0.5 MPa), but spatial and temporal variability large; using a single average introduces significant error	Weyburn measured pore pressure; Phase 1A reports

Table 2. *Cont.*

Parameter	Value/Range	Unit	Uncertainty Range	Data Source
Temperature (T)	Phase 1A average ≈ 63 °C; experiments commonly 59–60 °C	°C	Well-specific error ±2–5 °C; spatial variation may be ±5–10 °C	Weyburn well temperature logs; regional geothermal gradient
Salinity (S)	Typical reservoir brine 35–110 g/L; experiments include ≈64, 72, 80, 115 g/L	g/L (or mol/kg)	Large spatial variation; single-well representative uncertainty ±15–20%	Weyburn formation water analyses; experimental data from the literature
CO_2 solubility (C_{CO2})	≈0.7–1.0 mol CO_2/kg H_2O at 50 °C, 14 MPa, 35–110 g/L; Li et al. measured Weyburn brine at 59 °C up to 29 MPa	mol/kg or kg CO_2/m^3 brine	Solubility model ±5–10%; mainly propagated from P, T, S uncertainties	Li et al. correlation; PHREEQC or Duan & Sun model
Porosity (φ)	Marly: 16–38%; Vuggy: 8–20%	Dimensionless (fraction)	Core-to-log scale difference ±10–15%	Weyburn core/log interpretations
Thickness (h)	Marly: 3–10 m; Vuggy: 8–22 m; total 18–29 m	m	Stratigraphic interval error ±15–25%	Weyburn geological model, well-control stratigraphy
Water saturation (S_w)	≈35% initial discovery-phase estimate	Dimensionless (fraction)	Initial ±10–15%; post-injection changes unknown, potentially higher	Weyburn initial water saturation estimate; log interpretation
Contact fraction ($f_{contact}$)	Not provided; engineering judgment or simulation required	Dimensionless	Highly uncertain, possibly ±50% or higher	Engineering estimate, history matching, or numerical simulation
Swept fraction (f_{swept})	Not provided; engineering judgment or simulation required	Dimensionless	Highly uncertain, possibly ±50% or higher	Engineering estimate, history matching, or numerical simulation
Residual gas saturation (S_{gr})	No public measured uniform endpoint for 77 wells; some simulation studies use 0.05 but explicitly a sandstone assumption, not a Weyburn measurement	Dimensionless (fraction)	Carbonate S_gr highly variable; typical ±10–20%; using an incorrect endpoint causes systematic bias	Must come from Weyburn carbonate core CO_2–brine displacement/relative-permeability hysteresis data, or a history-matched model constrained by production, pressure, and 4D seismic
CO_2 density (ρ_{CO2})	Calculated from P, T	kg/m^3	Propagated from P, T uncertainties, about ±3–5%	NIST/REFPROP or PR/EOS equations
CO_2 purity (x_{CO2})	Not explicitly reported in current dataset	Mole fraction	Purity fluctuations possibly ±2–5%	Injection gas composition analysis

Table 2. *Cont.*

Parameter	Value/Range	Unit	Uncertainty Range	Data Source
Injection mass rate ($\dot{m}$)	Project-level ≈8500 t/d CO_2 supply/injection scale	t/d or kg/d	Metering error ±5–10%; project-level data not representative of individual wells	Public project-level data
Recovered CO_2 mass (M_{rec})	Possibly 0 or from production/recycling records	kg	If ignored, may overestimate net injection; depends on record completeness	Production/recycling records

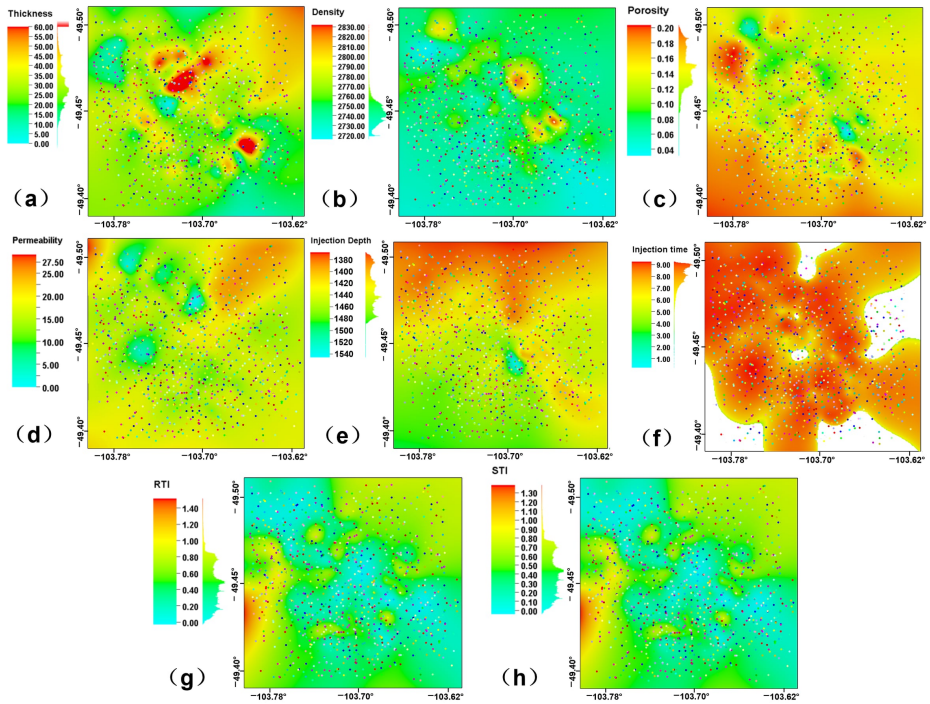


Figure 8. Map-view distribution of well-averaged input and output variables. Colors indicate value ranges for (**a**) thickness, (**b**) density, (**c**) porosity, (**d**) permeability, (**e**) injection depth, (**f**) injection time, (**g**) RTI, and (**h**) STI.

Table 3. Descriptive statistics of input and output variables used for model training and evaluation.

Type	Symbol	Unit	Minimum	Maximum	Mean	SD	Variance
Output	RTI	–	0.0161	0.979	0.408	0.239	0.0571
	STI	–	0.000362	0.887	0.227	0.21	0.0441
Input geological	Porosity	fraction	0.03	0.2	0.15	0.03	0.0009
	log10(K)	mD	−2	1.464	0.453	0.623	0.388
	Thickness	m	0.5	79.1	29.77	11	121
	Density	kg/m^3	2716	2832	2744	13.76	189.34
Input operational	Injection depth	m	1200	1800	1476	150	22,500
	Injection time	years	0.23	9.26	7.9	1.06	1.1236

Note. Thickness is the total reservoir thickness of the completed interval per well; the 18–29 m interval in Table 2 refers to the literature stratigraphic thickness of the Marly and Vuggy units, and compiled intervals exceed it locally. Porosity is the compiled dataset range from core and log interpretations and lies largely within the Vuggy range of Table 2. Injection depth is reservoir TVD.

4.2. Comparative Performance of Machine Learning Models

The six algorithms show differences in their exploratory STI predictions on the held-out subset (Figure 9). The R^2 column in Table 4 has been reconciled with the values annotated in Figure 9: XGBoost (0.81), LightGBM (0.76), Random Forest (0.71), KNN (0.60), SVM (0.56), and ElasticNet (0.53). The comparative ordering is therefore based on a single internally consistent set of displayed R^2 values. The models reported here were retrained after the depth predictor was changed from measured depth to the true vertical depth of the completed interval. The retraining did not change the displayed held-out R^2 values at two decimals: across the 77 wells, the MD-based and TVD-based depth values are almost rank-equivalent (Spearman $\rho = 0.98$), and the depth range collapses to the narrow reservoir interval of 1200–1800 m, so the splits and neighborhood structure of the fitted models are effectively unchanged. The retraining therefore changed the predictor definition rather than the reported performance.

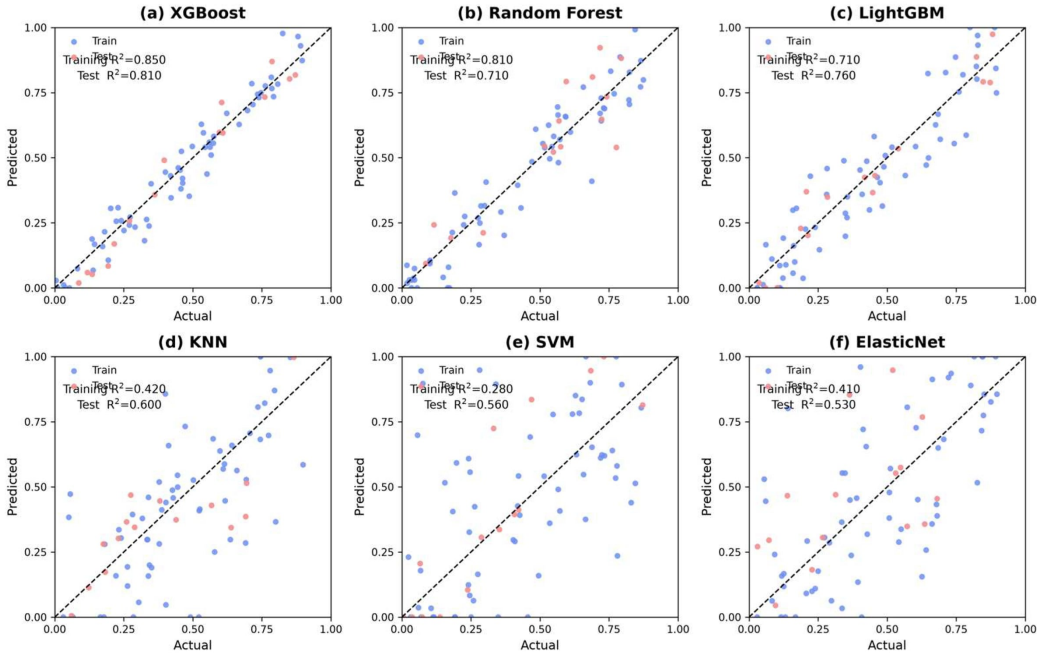


Figure 9. Predicted versus observed STI values for the six machine learning models on the held-out subset. Panels show (**a**) XGBoost, (**b**) Random Forest, (**c**) LightGBM, (**d**) KNN, (**e**) SVM, and (**f**) ElasticNet. The black dashed diagonal line denotes perfect agreement; the panel R^2 annotations correspond to the R^2 values in Table 4.

The strong XGBoost performance can be attributed to its sequential boosting strategy, regularization, and capacity to represent interactions among heterogeneous predictors. In this application, storage efficiency is influenced by the combined effects of pressure-temperature conditions, permeability architecture, pore volume, and injection duration. A boosted tree model can capture these interactions without assuming a single linear relationship. XGBoost was therefore selected as the champion model for the subsequent feature-importance analysis.

The sample-to-feature ratio is approximately 13:1, which is acceptable for a compact regression problem but still requires caution. Five-fold cross-validation within the training subset gave a mean cross-validated R^2 of 0.85, whereas the independent hold-out test subset gave $R^2 = 0.81$; the two quantities are reported here with their respective partitions to avoid ambiguity. This result, together with the hold-out test evaluation and regularized model settings, suggests that the model captures meaningful relationships rather than simply memorizing the training data.

The R^2 values reported in Table 4 and annotated in Figure 9 refer to the same displayed held-out comparison. The earlier mismatch caused by non-identical reporting outputs has been removed. No severe multicollinearity was detected; all VIF values were below 10. Pearson and Spearman correlations showed moderate associations between injection depth and some geological variables, but the overall structure did not indicate harmful redundancy. The reduced axes of Figure 9 now span the full observed and predicted STI range of zero to one.

Table 4. Exploratory held-out prediction performance of the six machine learning models for STI. The R^2 values are identical to the annotations in Figure 9.

Model	Original Permeability			$\log_{10}(K)$		
	R^2	MSE	MAE	R^2	MSE	MAE
XGBoost	0.81	0.021	0.1047	0.83	0.019	0.098
Random Forest	0.71	0.026	0.1014	0.72	0.025	0.1
LightGBM	0.76	0.023	0.112	0.77	0.022	0.11
KNN	0.6	0.034	0.0955	0.61	0.033	0.094
SVM	0.56	0.039	0.112	0.57	0.038	0.111
ElasticNet	0.53	0.047	0.1207	0.54	0.046	0.119

To further examine whether the model performance was sensitive to data partitioning, the prediction procedure was repeated under 100 independent random train–test splits. For the XGBoost model, the repeated-splitting analysis yielded a mean test-set R^2 of 0.810 ± 0.0280, with an empirical 95% confidence interval of $[0.7550, 0.8650]$. The corresponding mean MSE was 0.0137 ± 0.0049, with a 95% confidence interval of $[0.0058, 0.0266]$. The mean MAE was 0.0926 ± 0.0174. These results were obtained from 100 repeated random 80:20 train–test splits of the 77-well Weyburn dataset. These results indicate that the prediction accuracy remained relatively stable across different random partitions of the dataset.

4.3. Controls on CO_2 Storage Efficiency

Feature-importance analysis of the optimized XGBoost model was used to convert prediction results into geological interpretation. The normalized importance values shown in Figure 10 indicate a ranked control system rather than a single dominant variable.

The reported feature importance is a global ranking from the fitted XGBoost model and is not a statistical significance test. Feature importance was evaluated using 100 repeated permutation importance runs on the held-out test data. For each predictor, the mean importance, standard deviation, empirical 95% confidence interval, and permutation-test *p*-value were calculated. The resulting ranking is reported in Figure 10, and the corresponding SHAP-based interpretation is shown in Figure 11.

(1) Exploratory ranking. Injection depth and permeability are the two highest-ranked variables in the displayed XGBoost importance chart. This ranking is an association within the present analytical index dataset and is not a verified causal hierarchy.
(2) Other ranked variables. Porosity, rock density, and injection duration occupy intermediate positions in the displayed model ranking, while gross thickness has the lowest relative importance. Their relative values are reported with the mean importance, standard deviation, empirical 95% confidence interval, and permutation-test *p*-value obtained from the 100 repeated permutation importance runs described in Section 4.3.

(3) Interpretation boundary. Gross thickness may be a weaker proxy than net connected thickness and permeability architecture, but this proposition requires independent geological validation and cannot be inferred from the global importance chart alone.

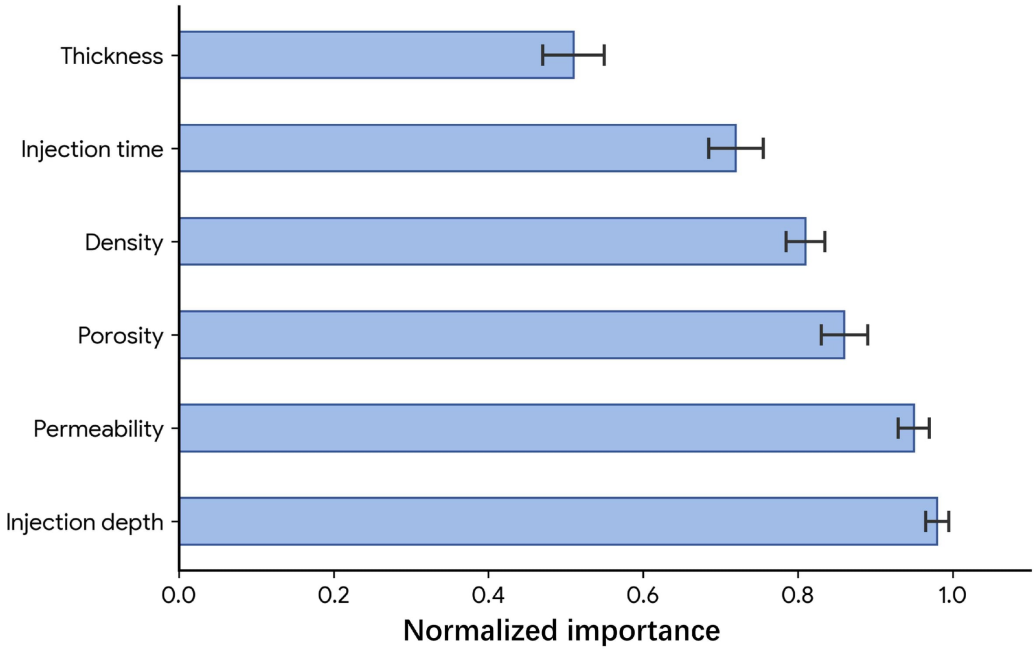


Figure 10. Feature importance derived from the optimized XGBoost model, calculated as the mean of 100 repeated permutation importance runs on the held-out test set. Error bars represent ±1 standard deviation across the 100 repetitions. Injection depth and permeability are the highest-ranked controls, followed by porosity, density, and injection duration; gross thickness contributes least strongly in this dataset.

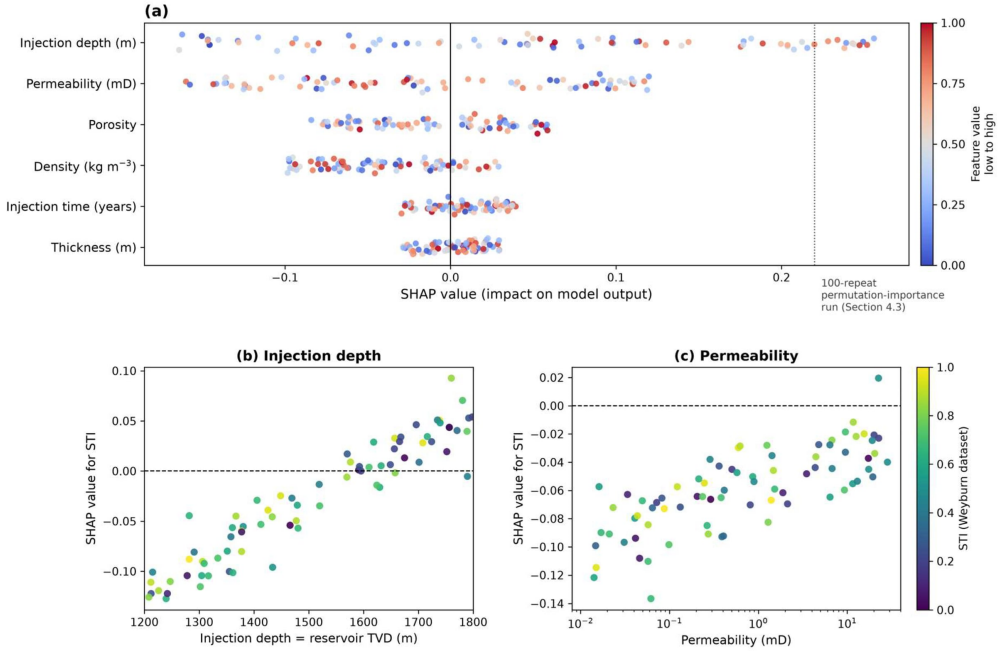


Figure 11. SHAP-based interpretation of the optimized XGBoost model for STI prediction. (**a**) SHAP analysis for the optimized XGBoost model on the 77-well Weyburn dataset. It shows the global distribution, magnitude, and direction of feature contributions; (**b**) SHAP dependence plot for injection depth; and (**c**) SHAP dependence plot for permeability. Each point represents one well, and color denotes the STI value.

The displayed ranking is used only to guide hypothesis generation for additional characterization. It does not establish that a single predictor is a dominant physical control across the Weyburn field.

SHAP-based interpretation of the optimized XGBoost model shows the global distribution, magnitude, and direction of feature contributions, with dependence plots for injection depth and permeability indicating systematic effects on predicted STI (Figure 11).

4.4. Practical Value of the Data-Driven Workflow

The integrated dataset and machine learning workflow developed here provide a rapid alternative for preliminary storage-efficiency screening. Once trained, the model can generate a preliminary STI estimate rapidly for wells or scenarios within the data-supported geological domain, without repeated full-physics simulation. However, this speed should not be interpreted as a guarantee of predictive accuracy away from sampled locations. The spatial block validation yielded an aggregated R^2 of 0.61, and the distance-buffer analysis reduced R^2 to 0.38 as the required train–test separation increased, with a repeated-split confidence interval that does not include zero. The workflow should therefore be used as a conditional screening aid for within-field interpolation and hypothesis generation, not as a validated predictor for unsampled areas or a substitute for site-specific simulation and monitoring.

The Weyburn case also demonstrates how a data-driven model can reproduce geologically plausible controls. The high importance of injection depth is consistent with the role of pressure and temperature in CO_2 phase behavior, whereas the high importance of permeability reflects the influence of carbonate heterogeneity on plume migration and trapping. With retraining on site-specific data, the workflow can be transferred to other carbonate CO_2-EOR and storage projects as a decision-support tool rather than a replacement for detailed reservoir simulation.

5. Discussion

5.1. Applicability of Data-Driven Storage Assessment

Analytical equations provide a transparent basis for calculating trapping indices, but they do not fully address the prediction problem faced during real CCUS screening. Field datasets are commonly incomplete, noisy, and collected at different scales, while reservoir behavior depends on interactions among geological and operational variables. Machine learning is useful in this setting because it can learn these interactions directly from integrated datasets and generate rapid predictions for many candidate cases [31,32].

The present results support three practical advantages of data-driven assessment. First, nonlinear algorithms can represent parameter interactions that are difficult to express in simplified screening equations. Second, trained models can evaluate many scenarios quickly, which is useful for sensitivity analysis and decision support. Third, model interpretation provides a ranked view of controlling variables, allowing prediction results to be translated back into geological reasoning rather than used as a purely black-box output.

5.2. Data-Representation Limits and Improvement Strategies

The main limitation of the current dataset is its well-level representation. Vertical and horizontal wells were reduced to single averaged values over completed intervals, which makes the model input consistent but compresses lateral and vertical heterogeneity. This simplification is particularly important for carbonate reservoirs, where flow may be concentrated in thin high-permeability zones, vugs, fractures, or stratigraphic layers [33–37]. Future datasets should retain interval-scale or trajectory-resolved descriptors where possible.

The model also omits several dynamic and thermodynamic variables because they were not consistently available, including time-varying injection rate, salinity, measured formation pressure, and temperature [38–42]. This omission limits the interpretation of injection depth, which is treated here as a proxy for pressure-temperature conditions rather than as a directly measured thermodynamic variable [43–47]. Future work should incorporate these controls and test time-series models, such as recurrent neural networks or transformer architectures, for forecasting trapping evolution. In addition, the present interpretation is based on feature-importance ranking rather than full explainability analysis; future studies should further analyze related methods to quantify local and global model behavior.

The correlation analysis showed that injection depth is moderately correlated with some geological parameters, which may reflect the combined influence of burial history and reservoir quality. Although VIF values did not indicate severe multicollinearity, the interpretation of depth importance should be cautious because depth is a proxy for pressure and temperature rather than a direct thermodynamic measurement. Future work should incorporate measured pressure–temperature data to decouple these effects.

5.3. Coupling Physical Understanding with Machine Learning

This study uses a hybrid formulation in which the RTI and STI are calculated from established analytical relationships and then predicted from geological and operational features. Although this approach is simpler than training directly on high-resolution three-dimensional simulation outputs, it provides a reproducible bridge between physical understanding and data-driven prediction. Because some predictors also enter the analytical trapping index framework, part of the learned signal may reflect the structure of the index calculation itself. The feature-importance ranking is therefore interpreted as a control ranking within the adopted screening index framework, not as a fully independent causal proof of storage physics. Future work should couple the workflow with high-fidelity multiphase-flow simulations and measured pressure–temperature–salinity data to reduce this circularity risk.

A natural next step is to incorporate outputs from high-fidelity multiphase-flow, geochemical, and geomechanical simulations into the training database. Such integration would allow the model to learn more complex storage behavior, including feedback among pressure buildup, brine displacement, dissolution, mineral reaction, and fault or fracture response. This direction would move the workflow toward a physics-informed machine learning framework for CCUS process evaluation.

5.4. Uncertainty Quantification and Validation Needs

Core, log, and injection-derived parameters contain measurement error, interpretation uncertainty, and scale-conversion uncertainty. These uncertainties were not propagated through the current modeling workflow [38–40]. Future studies should incorporate uncertainty propagation using Monte Carlo sampling, noise injection, or probabilistic input distributions.

Here, a conventional random 80:20 split may be overly optimistic for the Weyburn dataset because nearby wells can share similar facies, petrophysical properties, completion conditions, and injection histories. The random-split results therefore primarily reflect interpolation performance within the sampled field domain and do not, by themselves, demonstrate spatial extrapolation capability.

To address this issue, we added a spatial block cross-validation analysis. Well locations were represented by the coordinates of the reservoir intersection or, for horizontal wells, the midpoint of the completed interval and projected to a metric coordinate system before spatial grouping. The 77 wells were partitioned into five geographically contiguous blocks using

k-means clustering on the projected well coordinates. In each outer fold, one spatial block was withheld for testing, and the remaining four blocks were used for model development. Feature scaling, logarithmic transformation of permeability, and hyperparameter tuning were performed exclusively within the training blocks.

For each outer spatial fold, prediction performance was evaluated using R^2, MSE, and MAE. The XGBoost model achieved an aggregated spatial-block-validation R^2 of 0.61, MSE of 0.038, and MAE of 0.128. Across the five held-out spatial blocks, the mean R^2 was 0.59 ± 0.17, compared with 0.81 under random splitting. This reduction in performance indicates that the random split was optimistic because of spatial dependence among nearby wells.

A distance-buffer sensitivity analysis is also conducted. For each held-out spatial block, training wells located within a minimum separation distance d from any test well were excluded. The Euclidean separation distance was calculated from projected coordinates as:

$$d_{ij} = \sqrt{\left(X_i - X_j\right)^2 + \left(Y_i - Y_j\right)^2}, \quad (17)$$

Prediction errors were evaluated at d = 0, 0.5a, a, and 1.5a, where a is the horizontal variogram range of the reservoir-property field. The horizontal variogram range of the reservoir-property field is a ≈ 1500 m, so the buffer distances correspond to approximately 750 m, 1500 m, and 2250 m. As the required train–test separation increased, R^2 changed from 0.81 to 0.38, while MSE changed from 0.021 to 0.062. These results quantify the decline in predictive accuracy when the model is applied farther from sampled wells.

The spatial structure is further assessed in the out-of-block residuals using Moran's I with a distance-based spatial-weight matrix. The residual Moran's I was 0.11 with permutation-test $p = 0.23$, indicating no statistically significant residual spatial autocorrelation after spatial validation.

5.5. Geological Meaning of the Feature-Importance Results

The feature-importance analysis is an exploratory model interpretation, not a direct demonstration of trapping mechanisms. Injection depth is retained as a well-location and operational descriptor. Published field-scale reference conditions cannot substitute for well-level measurements of pressure, temperature, salinity, or phase state; depth is therefore not treated as a measured thermodynamic proxy in this dataset.

Permeability may be related to flow connectivity, vugs, fractures, and diagenetic fabric in carbonate reservoirs. However, its plotted model importance is treated as an association within the analytical index framework, and the severely skewed distribution was therefore evaluated through the documented log-transformation sensitivity analysis reported in Table 4, which supports the stability of the ranking.

Porosity and rock density form a second tier of controls. Porosity sets the volume available for CO_2 storage, but its effect depends on whether pores are connected and accessible to flow. Density can indirectly reflect lithology, mineral composition, and cementation, all of which may influence pore structure and reactive surface area. Although mineral trapping is not explicitly modeled in this workflow, these relationships help explain why density contributes to the prediction. Because the compiled porosity range of 0.03–0.20 lies largely within the Vuggy interval reported in Table 2, the dataset is dominated by Vuggy-type pore systems, and the porosity control inferred here should be read in that context rather than extrapolated to the high-porosity Marly intervals.

The low importance of gross thickness is geologically meaningful. In a stratified reservoir, storage efficiency is governed less by total stratigraphic thickness than by the net thickness, continuity, and permeability of connected flow units. Internal baffles between the Marly and Vuggy intervals can compartmentalize flow, so a thicker interval does not necessarily provide a proportionally larger effective storage volume.

Injection duration shows a moderate influence, reflecting the time dependence of dissolution and residual trapping. Longer injection periods generally increase the opportunity for CO_2–brine contact and plume immobilization, but the rate of this process depends strongly on permeability architecture. The combined ranking therefore supports a conceptual model in which thermodynamic conditions and flow connectivity jointly control storage efficiency in the Weyburn field.

6. Conclusions

This study developed and evaluated a feature-importance-supported machine learning workflow for estimating STI-based CO_2 storage efficiency in the Weyburn carbonate reservoir. By integrating geological and operational information from 77 wells, the analysis suggests that data-driven models can capture nonlinear relationships between reservoir heterogeneity and analytical trapping indicators while still requiring caution because of the compact sample size and analytical index target construction. The main conclusions are as follows:

In the displayed held-out comparison, XGBoost has the highest R^2 value (0.81). This exploratory result does not establish robust superiority because repeated, nested, and spatially blocked validation results have been performed using a traceable final dataset.

The displayed global ranking places injection depth and permeability ahead of the other input variables. This ordering is treated as an exploratory association within the adopted analytical screening index framework, not as a verified thermodynamic or causal conclusion.

The relatively low importance of gross thickness shows that net connected storage volume, internal baffles, and vertical permeability distribution are more informative than total stratigraphic thickness in the Weyburn system. This finding supports high-resolution geological characterization during storage screening.

The workflow provides a transferable screening approach for carbonate CO_2 storage projects, but transferability should be regarded as conditional rather than automatic. Because the log-transformation sensitivity test and the spatially aware validation have now been completed (Table 4; Section 5.4), the remaining development needs are the incorporation of site-specific salinity, measured pressure-temperature data, injection-rate history, and uncertainty analysis, and the framework can support faster and more transparent evaluation of geological storage potential.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/pr14213532/s1.

Author Contributions: Conceptualization, C.J.; methodology, C.J.; validation, G.H.; formal analysis, G.H.; writing—original draft preparation, C.J.; writing—review and editing, G.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the CNPC Liaohe Oilfield Project (2023KJZX-02).

Data Availability Statement: The 77-well table of the six predictors with the corresponding STI and RTI values and the script computing STI and RTI from Equations (1)–(7) have been deposited in Table S1.

Acknowledgments: The authors acknowledge the data and technical resources that supported compilation of the Weyburn well-level dataset.

Conflicts of Interest: Author Caihua Jia was employed by the China National Petroleum Corporation Liaohe Oilfield Company Oil Production Technology Research Institute. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The China National Petroleum Corporation Liaohe Oilfield Company Oil Production Technology Research Institute had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Babarinde, O.; Okwen, R.; Frailey, S.; Yang, F.; Whittaker, S.; Sweet, D. A workflow to assess the efficacy of brine extraction for managing injection-induced seismicity potential using data from a CO_2 injection site near Decatur, Illinois. *Int. J. Greenh. Gas. Control* **2021**, *109*, 103393. [CrossRef]
2. Cheng, Y.; Liu, W.; Xu, T.; Zhang, Y.; Zhang, X.; Xing, Y.; Feng, B.; Xia, Y. Seismicity induced by geological CO_2 storage: A review. *Earth-Sci. Rev.* **2023**, *239*, 104369. [CrossRef]
3. Al-Sakkari, E.G.; Ragab, A.; Dagdougui, H.; Boffito, D.C.; Amazouz, M. Carbon capture, utilization and sequestration systems design and operation optimization: Assessment and perspectives of artificial intelligence opportunities. *Sci. Total Environ.* **2024**, *917*, 170085. [CrossRef] [PubMed]
4. Ma, J.; Han, Y.; Wang, M.; Zhong, W.; Du, W.; Qian, F. Perspective on artificial intelligence for carbon capture utilization and storage (CCUS) in Petrochemical Industry. *Carbon Capture Sci. Technol.* **2025**, *16*, 100471. [CrossRef]
5. Park, J.Y.; Kihm, J.H. Numerical simulation of probabilistic CO_2 behavior based on 3D lithology model in heterogeneous geologic formation in the Janggi Basin, Korea. *Geosci. J.* **2025**, *29*, 788–801. [CrossRef]
6. Nait Amar, M. Towards improved genetic programming based-correlations for predicting the interfacial tension of the systems pure/impure CO_2-brine. *J. Taiwan Inst. Chem. Eng.* **2021**, *127*, 186–196. [CrossRef]
7. Zendehboudi, S.; Seyyedattar, M.; Cata Saady, N.M.; Mohammadzadeh, O.; Mamudu, A.; Heagle, D. Exploring the Potential of Offshore Geological CO_2 Storage in Canada: A Comprehensive Review and Future Outlook. *Energy Fuels* **2025**, *39*, 5987–6025. [CrossRef]
8. Bachu, S. Screening and ranking of sedimentary basins for sequestration of CO_2 in geological media in response to climate change. *Environ. Geol.* **2003**, *44*, 277–289. [CrossRef]
9. Ramírez, A.; Hagedoorn, S.; Kramers, L.; Wildenborg, T.; Hendriks, C. Screening CO_2 storage options in The Netherlands. *Int. J. Greenh. Gas. Control* **2010**, *4*, 367–380. [CrossRef]
10. Sanguinito, S.; Goodman, A.L.; Sams, J.I. CO_2-SCREEN tool: Application to the Oriskany Sandstone to estimate prospective CO_2 storage resource. *Int. J. Greenh. Gas. Control* **2018**, *75*, 180–188. [CrossRef]
11. Hsu, C.-W.; Chen, L.-T.; Hu, A.H.; Chang, Y.-M. Site selection for carbon dioxide geological storage using analytic network process. *Sep. Purif. Technol.* **2012**, *94*, 146–153. [CrossRef]
12. Alcalde, J.; Heinemann, N.; James, A.; Bond, C.E.; Ghanbari, S.; Mackay, E.J.; Haszeldine, R.S.; Faulkner, D.R.; Worden, R.H.; Allen, M.J. A criteria-driven approach to the CO_2 storage site selection of East Mey for the Acorn project in the North Sea. *Mar. Pet. Geol.* **2021**, *133*, 105309. [CrossRef]
13. Stauffer, P.H.; Viswanathan, H.S.; Pawar, R.J.; Guthrie, G.D. A system model for geologic sequestration of carbon dioxide. *Environ. Sci. Technol.* **2009**, *43*, 565–570. [CrossRef] [PubMed]
14. Rohmer, J.; Bouc, O. A response surface methodology to address uncertainties in cap rock failure assessment for CO_2 geological storage in deep aquifers. *Int. J. Greenh. Gas. Control* **2010**, *4*, 198–208. [CrossRef]
15. Pan, F.; McPherson, B.J.; Dai, Z.; Jia, W.; Lee, S.-Y.; Ampomah, W.; Viswanathan, H.; Esser, R. Uncertainty analysis of carbon sequestration in an active CO_2-EOR field. *Int. J. Greenh. Gas. Control* **2016**, *51*, 18–28. [CrossRef]
16. IEEFA. *Carbon Capture in the Southeast Asian Market Context*; IEEFA: Valley City, OH, USA, 2022.
17. Mortezaei, K.; Vahedifard, F. Multi-scale simulation of thermal pressurization of fault fluid under CO_2 injection for storage and utilization purposes. *Int. J. Rock Mech. Min. Sci.* **2017**, *98*, 111–120. [CrossRef]
18. Verdon, J.P. Using microseismic data recorded at the Weyburn CCS-EOR site to assess the likelihood of induced seismic activity. *Int. J. Greenh. Gas. Control* **2016**, *54*, 421–428. [CrossRef]
19. Xie, C.; Du, S.; Wang, J.; Lao, J.; Song, H. Intelligent modeling with physics-informed machine learning for petroleum engineering problems. *Adv. Geo-Energy Res.* **2023**, *8*, 71–75. [CrossRef]
20. Laloy, E.; Hérault, R.; Lee, J.; Jacques, D.; Linde, N. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. *Adv. Water Resour.* **2017**, *110*, 387–405. [CrossRef]
21. Alhajeri, M.S.; Abdullah, F.; Wu, Z.; Christofides, P.D. Physics-informed machine learning modeling for predictive control using noisy data. *Chem. Eng. Res. Des.* **2022**, *186*, 34–49. [CrossRef]
22. Hui, G.; Chen, S.; He, Y.; Wang, H.; Gu, F. Machine learning-based production forecast for shale gas in unconventional reservoirs via integration of geological and operational factors. *J. Nat. Gas. Sci. Eng.* **2021**, *94*, 104045. [CrossRef]
23. Koray, A.-M.; Bui, D.; Kubi, E.A.; Ampomah, W.; Amosu, A. Machine learning based reservoir characterization and numerical modeling from integrated well log and core data. *Geoenergy Sci. Eng.* **2024**, *243*, 213296. [CrossRef]
24. Cao, L.; Jiang, F.; Chen, Z.; Gao, Y.; Huo, L.; Chen, D. Data-driven interpretable machine learning for prediction of porosity and permeability of tight sandstone reservoir. *Adv. Geo-Energy Res.* **2025**, *16*, 21–35. [CrossRef]
25. Hui, G.; Pi, Z.; Ge, C.; Li, Y.; Chen, Z.; Wu, D.; Lu, Y.; Ni, F.; Yang, X.; Zhang, Y.; et al. Managing Hydraulic Fracturing-Induced Seismicity in the Duvernay Shale: Triggering Mechanisms, Influencing Factors, and Practice-Oriented Mitigation Strategies. *Int. J. Coal Sci. Technol.* **2026**, *13*, 85. [CrossRef]

26. Chen, F.; Sun, L.; Jiang, B.; Huo, X.; Pan, X.; Feng, C.; Zhang, Z. A Review of AI Applications in Unconventional Oil and Gas Exploration and Development. *Energies* **2025**, *18*, 391. [CrossRef]
27. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [CrossRef]
28. Bui, T.V.; Dao, H.H.; Nguyen, H.T.; Ta, Q.D.; Le, H.N.N.; Kieu, P.; Mai, C.L.; Tran, T.D.; Nguyen, H.S.; Nguyen, H.D.; et al. A Critical Review on the Opportunities and Challenges of Offshore Carbon Capture, Utilization, and Storage. *Sustainability* **2025**, *17*, 9250. [CrossRef]
29. Ge, C.; Yang, X.; Hui, G.; Chen, Z.; Zhang, Y.; Pi, Z.; Li, Y.; Bao, P.; Wu, D.; Lu, Y.; et al. Triggering mechanisms and mitigation strategies of CO_2 injection-induced seismicity in the Canada Weyburn field. In *Proceedings of the SPE Asia Pacific CCUS Conference, Kuala Lumpur, Malaysia*; SPE: Richardson, TX, USA, 2025. [CrossRef]
30. White, D. Seismic Characterization and Time-Lapse Imaging during Seven Years of CO_2 Flood in the Weyburn Field, Saskatchewan, Canada. *Int. J. Greenh. Gas. Control* **2013**, *16*, S78–S94. [CrossRef]
31. Wu, X.; Xing, H.; Hu, Z.; Huang, Y.; Liu, M.; Lv, X.; Li, Q.; Yang, L.; Song, Y. Data-Driven Framework for Fast Screening and Multiobjective Optimization of CO_2 Storage: Learning Temporal Evolution from Global Case Studies. *Energy Fuels* **2026**, *40*, 6983–6994. [CrossRef]
32. Du, S.; Wang, S.; Zhang, Q. High-fidelity spatiotemporal evolution of CO_2 storage in saline aquifers: A physics-synergized deep learning framework for CO_2 plume prediction and super-resolution reconstruction. *Int. J. Coal Sci. Technol.* **2026**, *13*, 90. [CrossRef]
33. Bao, P.; Hui, G.; Hu, Y.; Song, R.; Chen, Z.; Zhang, K.; Pi, Z.; Li, Y.; Ge, C.; Yao, F.; et al. Comprehensive characterization of hydraulic fracture propagations and prevention of pre-existing fault failure in Duvernay shale reservoirs. *Eng. Fail. Anal.* **2025**, *173*, 109461. [CrossRef]
34. Cheng, C.; Chen, Z.-X.; Hui, G. Energy-Efficient Fracturing Based on Stress-Coupled Perforation. *Pet. Sci.* **2026**, *23*, 4873–4890. [CrossRef]
35. Nassabeh, M.; You, Z.; Keshavarz, A.; Iglauer, S. Sub-surface Geospatial Intelligence in Carbon Capture, Utilization and Storage: A Machine Learning Approach for Offshore Storage Site Selection. *Energy* **2024**, *305*, 132086. [CrossRef]
36. Qiu, L.; Yang, Y.; Sai, Y.; Cheng, Y. Physics-Constrained Meta-Embedded Neural Network for Bottom-Hole Pressure Prediction in Radial Oil Flow Reservoirs. *Processes* **2026**, *14*, 89. [CrossRef]
37. Cao, Y.; Liang, Z.; Che, M.; Luo, J.; Sun, Y. Advanced Deep Learning Networks for CO_2 Trapping Analysis in Geological Reservoirs. *Sustainability* **2025**, *17*, 7359. [CrossRef]
38. Mansour Pour, K.; Voskov, D. Physics-Informed Neural Networks Based on Sequential Training for CO_2 Utilization and Storage in Subsurface Reservoir. *J. Mach. Learn. Model. Comput.* **2023**, *4*, 27–40. [CrossRef]
39. Wen, H.; Wu, J.; Zhu, Y.; Xing, X.; Wu, G.; Zhang, S.; Xian, C.; Li, N.; Xiao, C.; Zhou, Y.; et al. Production Prediction Method for Deep Coalbed Fractured Wells Based on Multi-Task Machine Learning Model with Attention Mechanism. *Processes* **2025**, *13*, 1787. [CrossRef]
40. Jia, L.; Shi, G.; Jiang, N.; Hu, S.; Li, G.; Zhang, Z. Artificial Intelligence-Based Oil Well Condition Diagnosis and Production Parameter Optimization. *Processes* **2025**, *13*, 3872. [CrossRef]
41. Xie, Q.; An, S.; Xie, H.; Wang, W.; Niasar, V. Pore-Scale Transition Behavior of Digital Carbonate Rock Dissolution During CO_2 Geo-Sequestration. *Water Resour. Res.* **2026**, *62*, 2. [CrossRef]
42. Zhang, Y.; Pi, Z.; Hui, G.; Zhang, R.; Ni, F.; Li, Y.; Ge, C.; Bao, P.; Zhang, Y.; Yang, X.; et al. Experimental Research on Surfactant Flooding and Design of Integrated Schemes for Energy Storage Fracturing and Production in Tight Sandstone Reservoirs. *Processes* **2025**, *13*, 4031. [CrossRef]
43. Mohammed-Sajed, O.K.; Rashid, F.; Glover, P.W.J.; Collier, R.E.; Lorinczi, P. Quantitative Diagenesis for the Characterization of CCUS Storage in Carbonates. *Energy Fuels* **2026**, *40*, 2703–2720. [CrossRef]
44. Huang, R.; Qiao, D.; Hui, G.; Liu, X.; Su, Q.; Wang, W.; Bi, J.; Ren, Y. A Fracture Extraction Method for Full-Diameter Core CT Images Based on Semantic Segmentation. *Processes* **2025**, *13*, 2221. [CrossRef]
45. Cheng, Y.; Luo, X.; Zhuo, Q.; Gong, Y.; Liang, L. Description of Pore Structure of Carbonate Reservoirs Based on Fractal Dimension. *Processes* **2024**, *12*, 825. [CrossRef]
46. Lin, W.; Zhao, X.; Li, M.; Zhuang, Y. Pore Structure Characterization and Fractal Characteristics of Tight Limestone Based on Low-Temperature Nitrogen Adsorption and Nuclear Magnetic Resonance. *Fractal Fract.* **2024**, *8*, 371. [CrossRef]
47. Wei, Z.; Cui, Y.; Su, Y.; Zhou, W. Fast History Matching and Flow Channel Identification for Polymer Flooding Reservoir with a Physics-Based Data-Driven Model. *Processes* **2025**, *13*, 2610. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.