frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 28 April 2025
DOI 10.3389/fenvs.2025.1562087

OPEN ACCESS

EDITED BY
Jill A. Engel-Cox,
University of Colorado Denver, United States

REVIEWED BY
Mohammad H. Elkady,
Texas A and M University, United States
Mostafa Saghafi,
Shahrood University of Technology, Iran
Derek Vikara,
National Energy Technology Laboratory (DOE), United States

*CORRESPONDENCE
Emil Attanasi,
attanasi@usgs.gov
Timothy Coburn,
tim.coburn@colostate.edu

RECEIVED 16 January 2025
ACCEPTED 11 April 2025
PUBLISHED 28 April 2025

CITATION
Attanasi E, Freeman P and Coburn T (2025) Machine learning provides reconnaissance-type estimates of carbon dioxide storage resources in oil and gas reservoirs. *Front. Environ. Sci.* 13:1562087. doi: 10.3389/fenvs.2025.1562087

Machine learning provides reconnaissance-type estimates of carbon dioxide storage resources in oil and gas reservoirs

Emil Attanasi[1]*, Philip Freeman[1] and Timothy Coburn[2]*

[1]U.S. Geological Survey, Geology, Energy and Minerals Science Center, Reston, VA, United States, [2]Department of Systems Engineering, Colorado State University, Fort Collins, CO, United States

Oil and gas reservoirs represent suitable containers to sequester carbon dioxide (CO_2) in a supercritical state because they are accessible, reservoir properties are known, and they previously contained stored buoyant fluids. However, planners must quantify the relative magnitude of the CO_2 storage resource in these reservoirs to formulate a comprehensive strategy for CO_2 mitigation. Even reconnaissance-type estimates of CO_2 storage resources of known oil and gas reservoirs may require complicated calculations involving 1) estimates of recoverable oil and gas, 2) reservoir properties (depth, temperature, pressure, etc.), and 3) the physical qualities of the retained fluids. We demonstrate the application of machine learning (ML) algorithms to bypass these computations to yield more rapid estimates of CO_2 storage resources in reservoirs capable of hosting CO_2 in a supercritical state. ML algorithms are computationally efficient because they do not impose the strong assumptions on the data-generating process that standard statistical or engineering procedures require. Further, ML algorithms can capture highly complex, particularly nonlinear, relationships among predictor variables. We demonstrate the application of four different ML algorithms using data from onshore and offshore oil and gas reservoirs in Europe, and show they perform well when predictions are compared to engineering estimates. The proposed methods and models provide an effective and novel way to more rapidly and directly determine the subsurface CO_2 storage capacity of oil and gas reservoirs around the world, information that operators, researchers, and policymakers alike require to meet energy transition and decarbonization goals.

KEYWORDS

machine learning, carbon dioxide storage resources, carbon dioxide sequestration, oil and gas reservoirs, supercritical carbon dioxide

1 Introduction

Carbon dioxide (CO_2) capture and geologic storage is one of the methods proposed to moderate the growth and even reduce CO_2 emissions to the atmosphere. To make meaningful contributions to this goal, potential geologic storage units must be identified, and the CO_2 storage resource assessed[1] (Bachu et al., 2007). Depleted oil and

1 This study uses the term CO_2 storage resource to denote pore space in the reservoir occupied by recoverable oil and gas that could be utilized to store CO_2. This terminology is consistent with Society of Petroleum Engineers (2017) accepted definitions

gas reservoirs[2] represent a class of such geologic structures that are already identified and whose storage potential may be appraised using known production and reservoir data. These reservoirs are appealing because data characterizing the reservoir and the contained hydrocarbons have already been collected by the producer and regulatory authority. Moreover, the structural integrity of these reservoirs has already been demonstrated because they have previously stored hydrocarbons.

In this study we develop machine learning (ML) approaches to make reconnaissance-type estimates of the CO_2 storage resources of known oil and gas reservoirs. These reconnaissance-type estimates are based on data that characterize the properties of individual oil and gas reservoirs located in a study area. Such data may be published in public or propriety databases. Reconnaissance estimates are particularly important to policymakers and planners because they can provide a comprehensive view of the potentially available resources required for environmental policy and planning decisions. The data-driven nature of ML allows such estimates to be computed more rapidly than those obtained through a workflow of conventional engineering calculations, and it permits nonlinearities in the data to be accommodated directly without complex computation.

The oil and gas reservoir data used here included 9,340 known oil and gas reservoirs located in onshore and offshore western Europe that are currently or were formerly producing (IHS Markit, 2021); now known as S&P Global, retrieved September 2021. The calculation of CO_2 storage resources for these reservoirs, even at the reconnaissance level, requires complex computations entailing estimates of recoverable oil and gas, reservoir properties, and the physical properties of the produced oil and gas (see Supplementary Appendix 1). To this end, we applied ML algorithms using reservoir and oil and gas properties to facilitate the calculation of the CO_2 storage resources. The algorithms are trained and tested using a screened subset of these data. This paper first provides a short literature review, then describes the reservoir data, and briefly identifies the ML algorithms applied along with their hyperparameters. The predictive performance of the algorithms is then compared and summarized, and final remarks and discussion are provided.

2 Literature review

Various engineering and volumetric methods have been used in recent years to estimate the geological resource that could be available to store captured carbon (Bradshaw et al., 2006; Bachu, 2008; Pingping et al., 2009; Popova et al., 2012; Goodman, 2012; Heidug, 2013; Gorecki et al., 2015; NETL, 2015; Cantucci et al., 2016; Goodman et al., 2016; Ajayi et al., 2019; Moore, 2022). U.S. Geological Survey (2013), Consoli (2016), Kearns et al. (2017), Sanguinito et al. (2020), Zhang et al. (2022), and others provide estimates of geological CO_2 storage capacity on a global or regional scale, and/or for specific depositional or operational settings. Peck et al. (2017), for example, specifically discuss best practices for quantifying CO_2 storage resources in the context of enhanced oil recovery (EOR). Argatan et al. (2018) address CO_2 storage in depleted oil and gas fields in the Gulf of Mexico, and Haagsma et al. (2020) and Jones et al. (2024) consider storage in Michigan Basin oil and gas reservoirs.

This body of literature suggests that the various estimates of geologic CO_2 storage resources are quite uncertain, a situation largely attributable to methodological differences and treatment of the subsurface physics. Kadeethum et al. (2023) note that the various subsurface reservoir environments in which CO_2 could be stored are rather heterogeneous in terms of their physical and flow characteristics, and the behavior of these phenomena are often nonlinear and difficult to model, requiring significant time and computational resources. On the other hand, ML provides a way to effectively capture the relationships among subsurface features, including any nonlinear behaviors, without having to specifically rely on the theoretical constructs and assumptions more often associated with engineering approaches.

In layman's terms, ML is "a method of data analysis that automates analytical model building" or "a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention" (SAS n. d.; Wilpon et al., 2017). Consequently, the use of ML has been suggested as an alternative, surrogate, or proxy approach with which to address various aspects of carbon capture, utilization, and storage (CCUS) (Dumakor-Dupey and Arya, 2021; Mahoob et al., 2022; Kadeethum et al., 2023). For example, You et al. (2019) use ML to evaluate EOR and CO_2 storage resources in the Farnsworth Unit in the Texas Panhandle. Smeenk and Leeuwenburgh (2023) and Ferreira et al. (2024) use deep learning approaches to evaluate CO_2 storage resources in depleted gas fields. Bakhshian et al. (2022) use deep learning to assess CO_2 storage resources in reservoirs containing residual gas. He et al. (2022) provide an ML workflow for CO_2 storage capacity in deep saline reservoirs. Chen et al. (2018), Chen and Pawar (2019), Thanh et al. (2020), and Abdulwarith et al. (2024) use ML to address CO_2 EOR and storage capacity in residual oil zones. Moreover, Wen et al. (2023) develop a nested Fourier neural network model trained on detailed reservoir model output data to produce high-resolution CO_2 storage resource estimates for site-specific reservoirs. Here, we consider a novel application of ML to formulate reconnaissance-type estimates of CO_2 geologic storage resources, with specific reference to western Europe.

3 Data description

Country-wide, regional, or basin-wide high-level or reconnaissance appraisals of the CO_2 storage resources of oil and gas reservoirs commonly assume CO_2 will replace the produced reservoir fluids in the reservoir's pore space (Bachu et al., 2007;

2 The U.S. Energy Information Administration (2000) defines conventional oil and gas reservoirs as underground formations containing an individual and separate pool (natural accumulation) of producible oil and/or gas that is confined by impermeable rock or water barriers and is characterized by a single natural pressure system

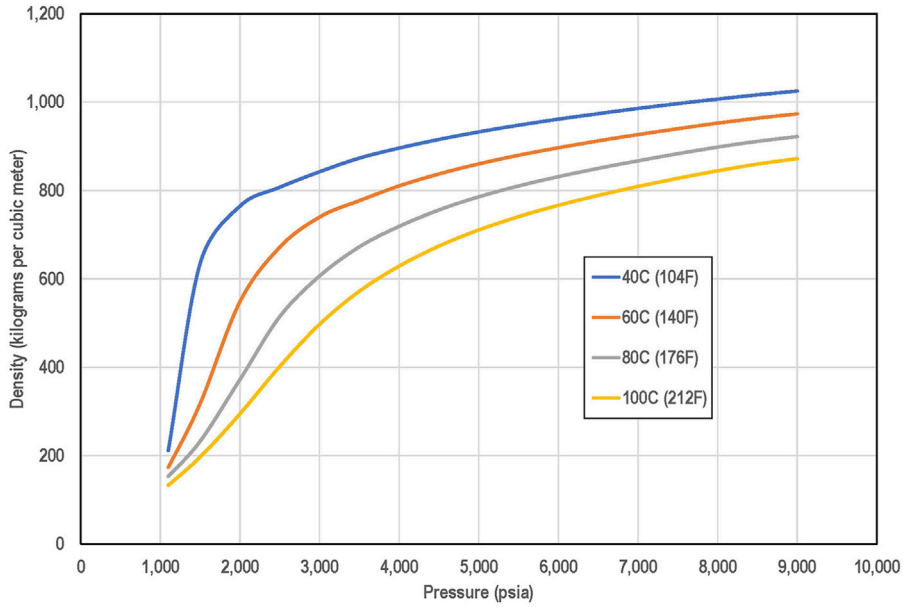

FIGURE 1
Density of supercritical carbon dioxide (CO_2) as a function of pressure at various temperatures. C is degrees Celsius, and F is degrees Fahrenheit. Calculations using the predictive model by Ouyang (2011) are based on the National Institutes Standards and Technology (NIST, 2024) data from Span and Wagner (1999).

Goodman et al., 2011; Cantucci et al., 2016; Aminu et al., 2017).[3] Because data on the location and appraisal of the CO_2 storage resources are critical for long-range planning for deployment of CO_2 transport and storage, the approach taken here is to base the CO_2 storage resource on an estimate of the reservoir volume of the recoverable hydrocarbons that can be expected to be produced from startup to abandonment. Supplementary Appendix 1 presents details of the multiple steps of the more conventional engineering procedure that uses the individual reservoir properties to estimate the CO_2 storage resource for each reservoir.

As of 2021, IHS Markit (2021) identified 9,340 oil and gas reservoirs located in western Europe. A reservoir was classified as oil if the natural gas to oil ratio, in terms of thousands of cubic feet (MCF) to barrels of oil (bbl), was less than 20 to 1 (Charpentier and Klett, 2005). Otherwise, it was classified as a non-associated gas reservoir (hereafter just called gas reservoir). Applying this criterion resulted in 3,943 oil reservoirs and 5,397 gas reservoirs. Total recoverable oil in the reservoirs is estimated to be 97.6 billion barrels and the total recoverable non-associated gas in the reservoirs is estimated to be 584 trillion cubic feet of gas (IHS Markit, 2021).

We first arrayed oil and gas reservoirs separately by the estimated subsurface volume of their contained hydrocarbons in order to screen by subsurface size. The largest 1,017 oil reservoirs by subsurface volume accounted for 95 percent of the cumulative subsurface oil reservoir volume, and similarly, the largest 1,619 gas reservoirs accounted for 95 percent of the cumulative gas reservoir volume. We further screened the oil and gas reservoirs using reservoir pressure and temperature data to identify reservoirs that could theoretically sustain injected CO_2 in its supercritical phase. That screening left 950 oil and 1,278 gas reservoirs. These 950 oil reservoirs accounted for 85.6 BBO of recoverable (producible) oil and the 1,278 gas reservoirs accounted for 509.1 TCF of producible gas.

A fundamental assumption of the procedure for computing the reconnaissance level estimates of CO_2 storage is that the volume of supercritical CO_2 that can be safely stored is equivalent to the reservoir volume of pore space vacated by the producible hydrocarbons. These estimates do not account for the injected CO_2 that might also be incidentally stored in formation water or in the residual oil and gas left in the reservoir after commercial production ceases. Alternatively, at a reconnaissance level, there is no way to determine which reservoirs might have been subject to subsidence after production, thus reducing the actual CO_2 storage resource.

We assumed that CO_2 is injected and stored in its supercritical state. Figure 1 shows how the density of supercritical CO_2 varies with pressures at various levels of constant temperature. For the pressure and temperature range in the figure, the supercritical CO_2 density increases as pressure increases and temperature declines.

The locations of individual reservoirs are proprietary to IHS Markit (2021), but cell maps show the sum total CO_2 storage resource of reservoirs within each 25-km square grid cell. Figure 2 is a map that shows the geographical distribution of the 950 oil reservoirs and Figure 3 shows the map of the 1,278 gas reservoirs used in this study. The total estimated CO_2 storage resource for the 950 screened oil reservoirs amounted to

3 Zhao et al. (2014) point out that injected CO_2 may also be retained in the formation water or the residual crude oil, or that it may be mineralogically retained in the reservoir rock formation. Furthermore, if extraction of the formation water is economic, the CO_2 storage resource can be expanded

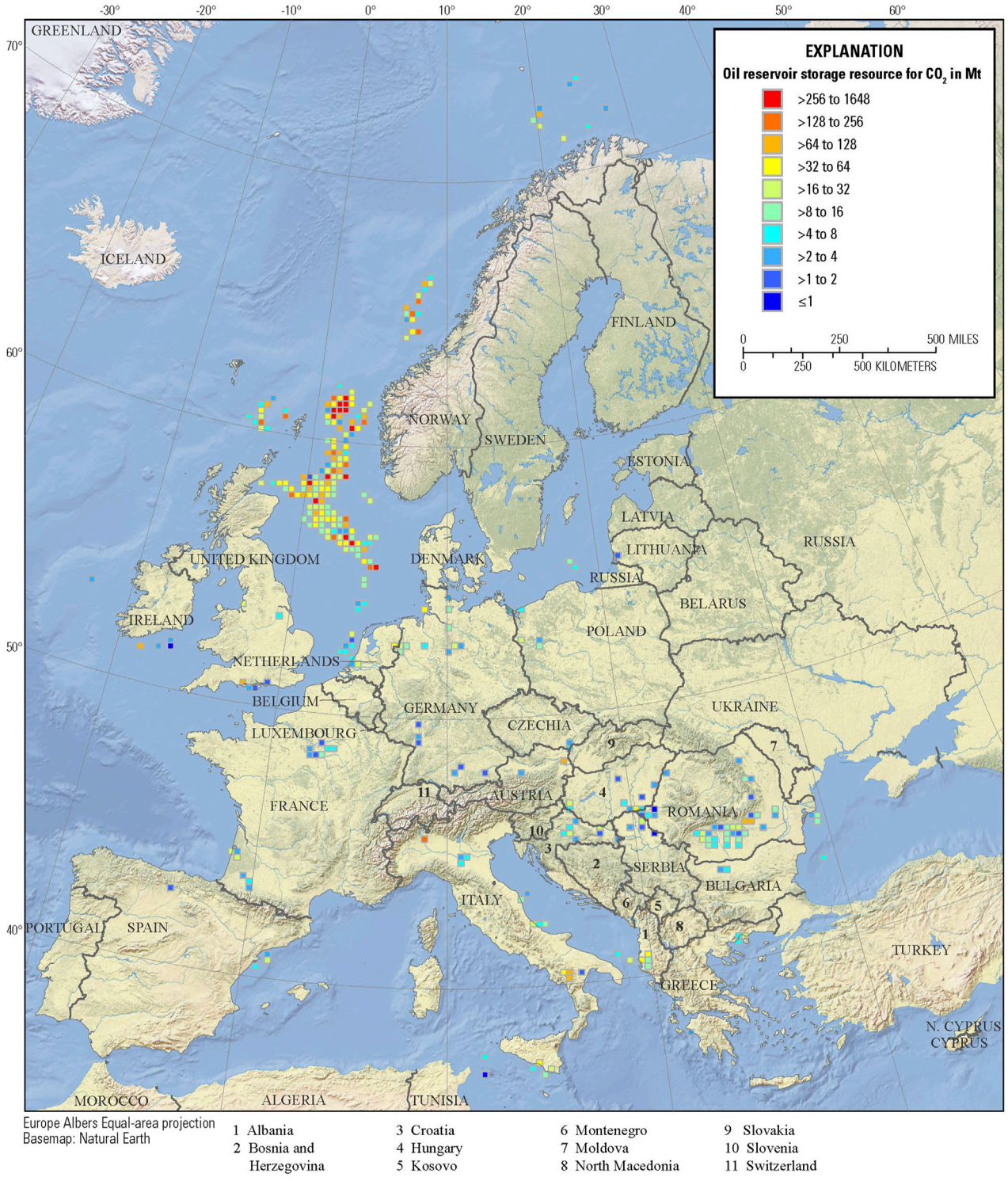

FIGURE 2
Estimated CO_2 storage resources in mass of megatons (Mt) for oil reservoirs in western Europe. Estimates of CO_2 storage resource are calculated from the estimates of volumes of recoverable oil from IHS Markit (2021). Cells of dimension 25 km by 25 km show the combined CO_2 storage resource of all individual oil reservoirs within the cell.

17.7 metric gigatons (GT), and similarly, the 1,278 CO_2 storage resources for the gas reservoirs amounted to 40.3 GT.

Both oil and gas reservoirs exhibit a right skewed size distribution in terms of recoverable hydrocarbon resources and estimated volumes of CO_2 storage resources. There are many small reservoirs which together account for a small fraction of the total resource in contrast to relatively few large reservoirs that account for most of the resource. Figures 4A,B show the

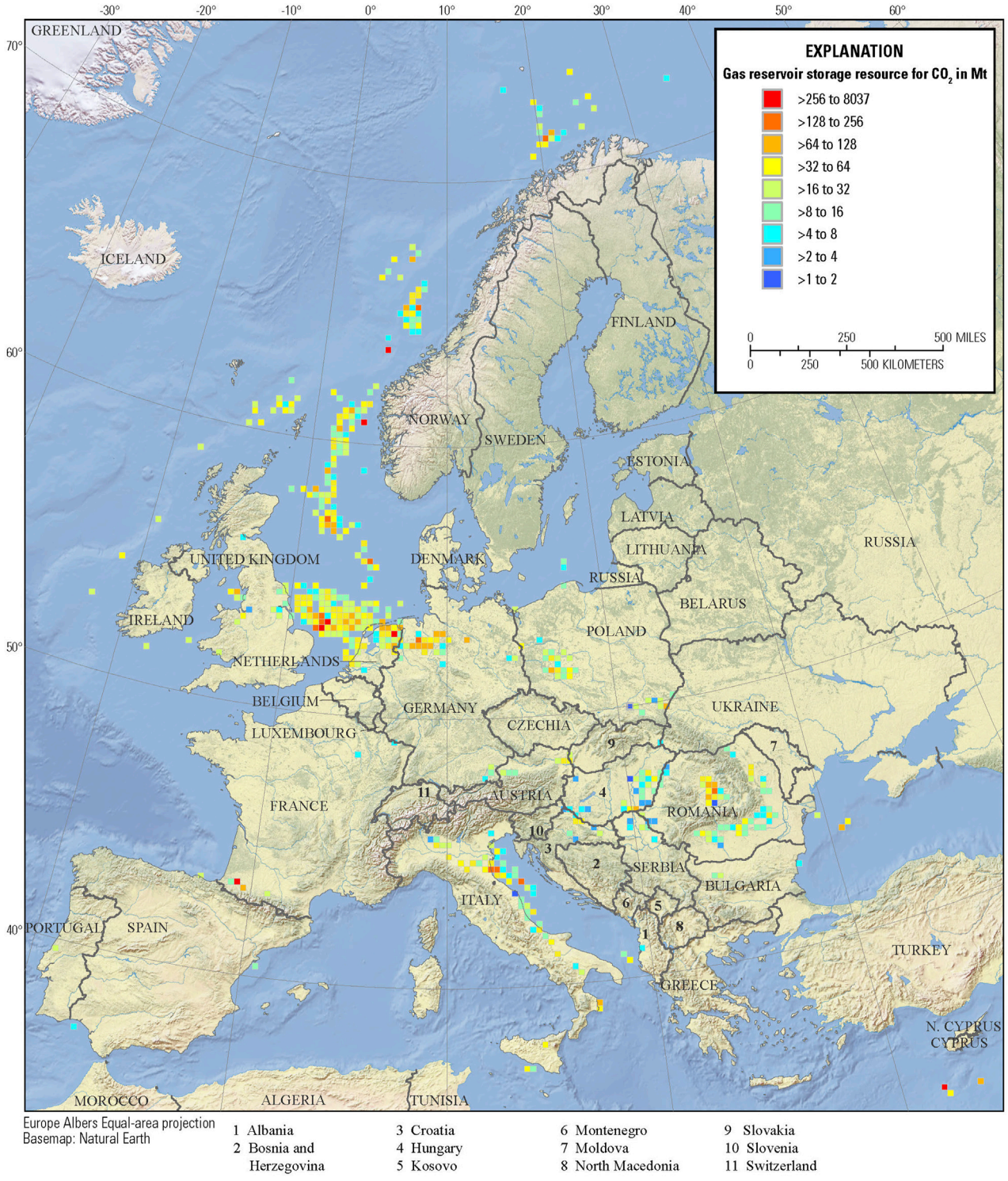

FIGURE 3
Estimated CO_2 storage resources in mass of megatons (Mt) for gas reservoirs in western Europe. Estimates of CO_2 storage resource are calculated from the estimates of volumes of recoverable gas from IHS Markit (2021). Cells of dimension 25 km by 25 km show the combined CO_2 storage resource of all individual gas reservoirs within the cell.

frequency volume distribution of the screened oil and gas reservoirs, respectively, and Figures 5A,B show the frequency volume distributions of the estimated CO_2 storage resource of the screened oil and gas reservoirs. In these figures, the horizontal axes providing the size classes are delineated in log base two in order to more clearly depict the wide range of sizes. In Figures 4A,B the largest five percent of the oil reservoirs (47) contain nearly half of the oil, and similarly, the largest 5 percent of the gas reservoirs (64)

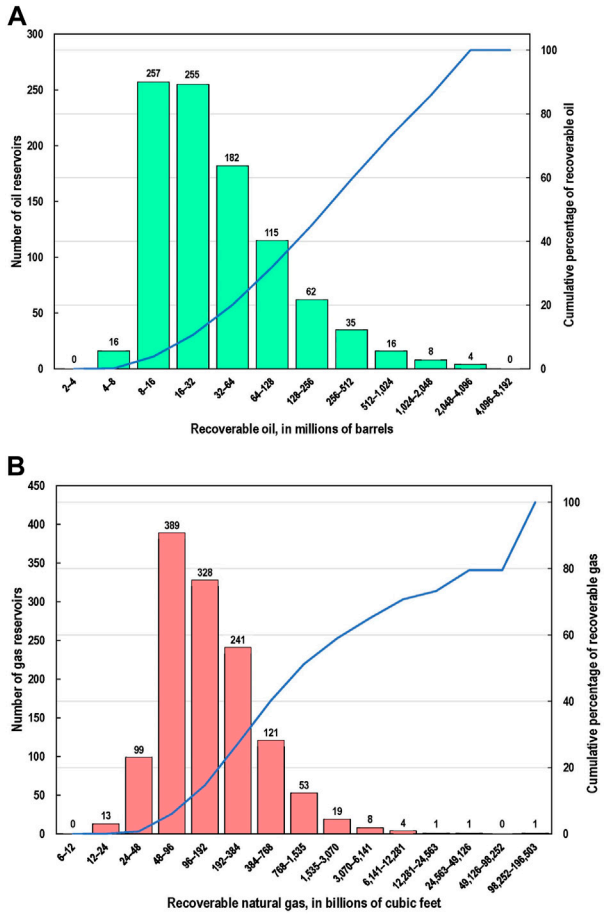

FIGURE 4
Distributions of reservoir sizes and respective cumulative percentages of recoverable resource volume: **(A)**, recoverable oil in millions of barrels in oil reservoirs, and **(B)**, recoverable gas in billions of cubic feet in gas reservoirs where the minimum size oil or gas reservoir is estimated to have the equivalent of 1 megaton (Mt) CO_2 storage resource. Note: The reservoir class size doubles on the x-axis.

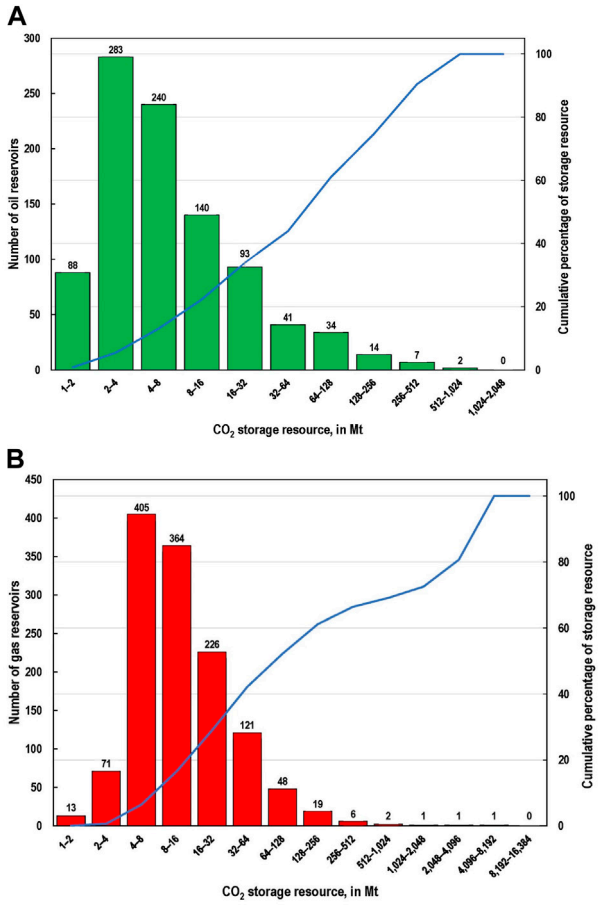

FIGURE 5
Distributions of CO_2 storage resource and cumulative percentage of total CO_2 storage resource volume: **(A)**, individual oil reservoir size and **(B)** individual gas reservoir size (minimum reservoir CO_2 storage resource is 1 megaton, Mt). The size class doubles on the x-axis.

contain 55 percent of the non-associated gas. Further, in Figures 5A,B the largest five percent of oil and gas reservoirs account for 53 percent and 55 percent of the estimated CO_2 storage resource, respectively.

4 Methodology

Oil and gas reservoirs were modeled separately because they use different measures for quality of the hydrocarbons, i.e., API gravity for oil and specific gravity for gas. Three observations of CO_2 storage resources beyond five standard deviations from the mean for the gas reservoirs were eliminated as extreme outliers, as well as one observation with an invalid gas specific gravity value. The final data sets represented 950 oil reservoirs and 1,274 gas reservoirs. Each reservoir dataset was then randomly divided into a training set containing 70 percent of the respective observations and a test set containing the remaining 30 percent. The predictor variables used for the CO_2 storage resource for oil reservoirs were reservoir temperature, reservoir pressure, API gravity, estimated oil recovery and estimated natural gas recovery. Predictor variables used for the CO_2 storage resource in gas reservoirs included reservoir temperature, reservoir pressure, specific gravity of the gas, and the estimate of the volume of recoverable gas from the reservoir.

Descriptive statistics of the predictor variables and CO_2 storage resource estimates are presented in Table 1, 2. The table shows the means and standard deviations, as well as the coefficients of variation (ratio of standard deviation to the mean) associated with each predictor variable in the training and test sets associated with the oil and gas reservoirs, respectively. Supplementary Appendix 2 provides box plots of the predictors and the CO_2 storage values for the oil and natural gas reservoirs (Supplementary Appendix 2; Figure 1), as well as tables of bivariate correlations (Supplementary Appendix 2; Tables 1, 2). The box plots for both the oil and gas reservoirs show observations concentrated in the smaller sizes of recoverable oil and gas, and for CO_2 storage resource. The bivariate correlation matrix shows the recoverable oil and gas having the strongest predictor correlations with the storage values of CO_2.

TABLE 1 Means, standard deviations, medians and coefficients of variation of predictor variables in the training and test sets of oil reservoirs [psia, pounds per square inch absolute; MMBO, million barrels oil; BCF, billion cubic feet; Mt, megatons or million metric tons; stddev, standard deviation; CV, coefficient of variation].

Set type	Number	Statistic	Temper-ature	Pressure	API gravity	Recover-able oil	Recoverable gas	CO_2 storage
			(°F)	(Psia)	(degree)	(MMBO)	(BCF)	(Mt)
Training set	665	mean	177.0	4,307.9	35.3	89.2	139.2	18.7
		stddev	54.3	2,133.6	6.7	260.4	836.3	60.5
		median	167.3	3,850.3	36.2	27.3	21.0	5.0
		CV	0.3	0.5	0.2	2.9	6	3.2
Test set	285	mean	174.6	4,183.8	34.4	92.5	88.9	18.5
		stddev	54.5	2,088.3	7.7	244.1	205.4	49.8
		median	168.6	3,773.0	36.0	28.4	20.0	5.4
		CV	0.3	0.5	0.2	2.6	2.3	2.7

TABLE 2 Means, standard deviations, medians and coefficients of variation of predictor variables in the training and test sets of gas reservoirs. [psia, pounds per square inch absolute; BCF, billion cubic feet; BCF, billion cubic feet; Mt, megatons or million metric tons; stddev, standard deviation; CV, coefficient of variation].

Set type	Number	Statistic	Temper-ature	Pressure	Specific gas gravity	Recoverable gas	CO_2 storage
			(°F)	(Psia)		(BCF)	(Mt)
Training set	891	mean	177.6	3,913.7	0.8	288.8	21.8
		stddev	59.1	1,708.4	0.1	607.1	41.6
		median	171.1	3,867.6	0.8	123.0	10.4
		CV	0.3	0.4	0.1	2.1	1.9
Test set	383	mean	179.7	3,935.2	0.8	267.2	22.0
		stddev	61.3	1,823.4	0.1	546.7	42.8
		median	173.0	3,847.2	0.8	116.5	10.2
		CV	0.3	0.5	0.1	2.0	1.9

Four different machine learning algorithms were applied to the data to construct predictive models. The algorithms and their respective hyperparameters, along with the procedures applied to the data are described in the following section. Each algorithm was trained on the same training dataset and the evaluation of predictive performance used the same test dataset.

4.1 Machine learning

4.1.1 Algorithms

Random Forest (RF) (Breiman, 2001), Gradient Boosting Trees (GBT) (Friedman, 2002), Extreme Boosting (XGBoost) (Chen and Guestrin, 2016), and deep neural network (DNN) (James et al., 2021) were used along with the values of the above-mentioned reservoir/resource predictor variables to directly model their relationship with the CO_2 storage resource estimates computed with the procedure outlined in Supplementary Appendix 1. The three tree-based algorithms, RF, GBT, and XGBoost, are related to the classification and regression tree (CART) algorithms of Breiman et al. (1984). Except for XGBoost, the computational routines are available from H2O.ai's open-source platform (H2O.ai, 2022). The XGBoost algorithm is published in R format (Chen et al., 2019). The DNN algorithm used here is also provided in the H2O platform, and it automatically standardizes the data prior to model calibration and then transforms the predictions back to the original metric (Candel and LeDell, 2023).

The tree-based approaches model non-linear relationships by partitioning the data space into progressively smaller groups so that data observations become more homogeneous within each partition. Individual trees are assembled by recursive partitioning that maximizes the local partition homogeneity with the tree structure being interpretable as a series of decision rules. The trees are assembled using the training data and predictive performance is evaluated by applying the trained model (algorithm) to the test dataset.

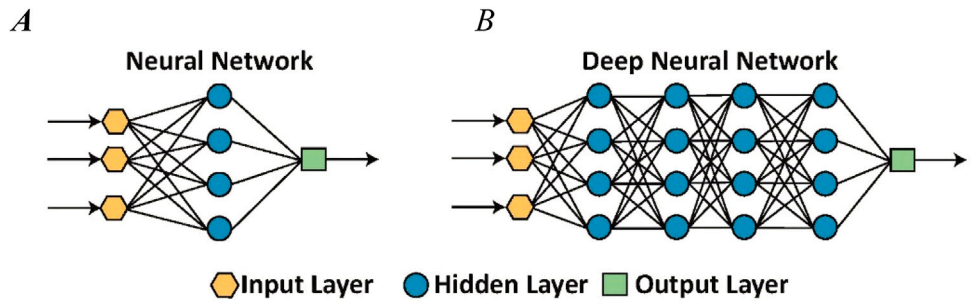

FIGURE 6
Illustration comparing schematics of **(A)** neural network and **(B)** deep neural network, DNN. Figure modified from Xing and Du (2019).

The RF algorithm generates an ensemble (or forest) of individual trees by applying bootstrap sampling to the training data and the predictors when forming individual trees. These sampling procedures are used to induce a degree of independence among individual trees. For regression analysis the predicted value of the target variable (in the present case, a reservoir's estimated CO_2 storage resource) is the average of the predicted values computed from all the individual trees in the forest (Breiman, 2001).

Individual trees are constructed sequentially by the GBT (Friedman, 2002) and XGBoost (Chen and Guestrin, 2016) algorithms rather than generating a forest of trees at the outset. The training data are used to construct a tree initially, but for the next iteration another tree is trained using the computed errors (residuals) associated with the predictions of the previous iteration. The "tree ensemble" is the sequence of fitted trees represented by the iterations, and target variable predictions are computed recursively using the constructed sequence of trees. Both the GBT and XGBoost algorithms entail bootstrap sampling of some fraction of each stage's residuals/training observations to fit successive trees and avoid overfitting. The XGBoost algorithm also samples the predictors at each iteration to build the next tree to further improve the robustness of predictions.

Hastie et al. (2009) describe single-layer artificial neural networks where predictor variable values are mapped using an activation to a single hidden layer composed of neurons (or activations). The hidden layer neurons, in turn, are mapped to a single-variable output layer by activation functions (Figure 6A).[4] These functions determine whether a neuron should be activated by computing the weighted function of input values (from the adjacent layer) and adding a bias term. The effect is to introduce non-linearity into the neuron output (James et al., 2021). Generally, any number of variables might be represented in both input and output layers.

A DNN network commonly has multiple hidden layers between the input and output layers. A schematic of a DNN having four hidden layers (layers of neurons between the input layer and the output layer) with each layer having four neurons is presented in Figure 6B. The network links represent unknown parameters (weights) that the network algorithm fits with the training data. Model predictions or outputs are calculated by forward propagation computations and the backward pass fits the neural network's unknown parameter weights (James et al., 2021). The architecture of the model, along with the specification of the activation functions, regularization parameters that include the dropout rate,[5] and the linear and quadratic penalty parameters, determines how well the DNN model fits the training data and how accurately it can predict the target variable values for the test data. (James et al., 2021).

4.1.2 Hyperparameters

Table 3 identifies each algorithm's hyperparameters and shows the parameter values used to train and test the models. For tree-based algorithms, hyperparameters place restrictions on the stochastic components, complexity, and size of the constructed trees. For DNN algorithms, hyperparameters restrict the size and the architecture of the neural network. The training dataset is used to calibrate the predictive model. The purpose of optimizing tuning hyperparameters is to minimize a measure of the expected error for new data (separate from the training dataset) when predicting the value of the target variable, which in the present case is the estimate of the reservoir's CO_2 storage resource. For all the algorithms evaluated here, we applied to the training data an estimate of the generalization error (e.g., the average root mean square error) obtained using a five-fold cross validation[6] approach. We generated the models' hyperparameters using an exhaustive grid search over a range of possible hyperparameter values (Halder, 2023). Table 3 in Supplementary Appendix 2 shows the range searched. Following the search, some adjustments were made to avoid overfitting the trained model.

4 An activation in a neural network is a mathematical function that collects and classifies information according to a specific architecture (number of layers, number of neurons).

5 The dropout rate randomly disengages certain nodes in a layer according to the stated dropout probability imposed during training

6 In cross-validation, the training set is divided into multiple folds, with one of the folds designated as a validation set, and the model is trained on the remaining folds. This process is repeated multiple times, and each time a different fold is used as the validation set. The performance measures from each of the validation folds are averaged to estimate the model's predictive performance when new data are applied

TABLE 3 Hyperparameter values specified for each model for oil reservoirs and for gas reservoirs.

Algorithm/Hyperparameter	Explanation	Oil reservoirs	Gas reservoirs
Random Forest			
Mtrie	Number of predictors sampled	3	3
Min_row	Minimum observations: terminal node	3	3
Max_depth	Maximum branches from root to leaf per tree	5	5
Ntrees	Number of trees	900	400
Sample_rate	Fraction of training data sampled	1	0.95
Gradient Boosting Trees			
Learn_rate	Weight contribution for each tree	0.05	0.05
Max_depth	Maximum branches from root to leaf per tree	2	4
Sample_rate	Fraction of training data sampled	0.9	1
Col_sample_rate	Fraction available data after sample_rate	0.9	0.9
Ntrees	Number of rounds	400	300
Extreme Boosting (XGBoost)			
Eta (learning rate)	Weight contribution for each tree	0.2	0.15
Max_depth	Maximum branches from root to leaf per tree	5	3
Minimum_child_weight	Threshold observation weights required for node	3	1
Subsample fraction	Fraction of training data sampled	1	1
Colsample_bytree	Fraction of predictors sampled	1	0.9
Ntrees	Number of rounds	77	53
Gamma	Fixed threshold of gain improvement for split	0	2
Lambda	Quadratic regularization parameter	0	0
Alfa	Linear regularization parameter	0	0
Deep Neural Network			
Hidden	Number of hidden layers	5	4
Neurons per hidden layer	Number for each hidden layer	15	20
Activation	Activation function choice depends on software	Rectifier[a]	Rectifier[a]
Input dropout rate	Dropout rate	0	0
Epochs	Iterations through training set	35	20
l1	Linear regularization parameter	0	0
l2	Quadratic regularization parameter	0	0

[a]Rectifiers are a type of activation function representing an option for the activation hyperparameter.

5 Predictive performance

Table 4 shows three measures of predictive performance. The first measure is the mean absolute error, the second is the root mean square error, and the third is the fraction of the variation in each dataset's target variable (predicted CO_2 storage resource) explained by the model. The mean absolute error is the average absolute difference between the predicted values and the actual target value, irrespective of direction. The root mean square error is the square root of the average of the squared differences between predicted values and actual target values, thus giving higher weight to large errors encountered in predicting extreme values. The variation explained is computed as one would compute the coefficient of determination (or R^2) commonly associated with regression analysis. The evaluation of predictive performance focusses on how well the trained models predict the target variable for the common test dataset.

TABLE 4 Predictive performance of four algorithms: Random Forest (RF), Gradient Boosting Trees (GBT), Extreme Boosting (XGBoost), and Deep Neural Network (DNN) [Mt, megatons or million metric tons].

Reservoir type	Training set				Test set			
	Number of observations	Mean absolute error (mt)	Root mean square (mt)	Variation explained	Number of observations	Mean absolute error (mt)	Root mean square (mt)	Variation explained
Random forest								
Oil reservoirs	665	4.4	16.5	0.926	285	5.2	15.6	0.903
Gas reservoirs	892	4.1	15.4	0.863	383	3.9	9.4	0.951
Gradient Boosting Trees								
Oil reservoirs	665	4.7	19.2	0.900	285	5.4	14.6	0.914
Gas reservoirs	892	3.3	13.0	0.901	383	3.7	13.7	0.897
Extreme Boosting (XGBoost)								
Oil reservoirs	665	0.8	1.5	0.999	285	4.0	17.7	0.873
Gas reservoirs	892	1.7	3.9	0.991	383	2.9	10.1	0.944
Deep Neural Network								
Oil reservoirs	665	3.0	7.3	0.985	285	3.8	12.0	0.942
Gas reservoirs	892	3.6	17.8	0.817	383	2.4	7.3	0.971

For three of the 4 ML algorithms, the predictive performance of the models associated with the gas reservoir CO_2 storage resources is superior to the corresponding models used to predict the CO_2 storage resource of oil reservoirs (see Table 3). Of the gas reservoir prediction models, the DNN model explained the target variable's variation (R^2) the best while having the lowest root mean square error. For the gas reservoir models, the tree-based algorithms were able to explain at least 90 percent of the target variable's variation for the test data. For the oil reservoirs, the DNN algorithm had superior predictive performance in terms of mean absolute error, root mean square error and explained variation statistics. The greater complexity of the DNN algorithm may account for its superior predictive performance for both oil and gas reservoirs[7].

Figures 7A,B show the predicted *versus* calculated (see Supplementary Appendix 1) CO_2 storage resources assigned to the test data oil reservoirs and gas reservoirs, respectively. The solid lines show cumulative storage resource growth across storage size classes, and the bars show storage volume attributed to each size class. The bar charts suggest that oil reservoir storage predictions at the size classes smaller than 16 million barrels substantially overestimate CO_2 storage values; and although the larger size classes are underestimated, the cumulative curve shows substantial displacement. The natural gas reservoir storage predictions substantially underestimate the calculated storage values in size classes greater than 32 Mt to 128 Mt, as well as for gas reservoirs with storage resource greater than 256 Mt. The graphs suggest why the root mean square errors associated with the test data are large even though the explained variation is relatively high.

Examination of the importance of the predictors may prove helpful in explaining the difference in predictive performance of both the oil and gas reservoir models. Table 5 shows the ordered relative importance (percent explanation) of the predictive variables for the RF, GBT, XGBoost, and DNN models. The leading predictors of CO_2 storage resources for oil reservoirs are the estimated recoverable oil and recoverable gas for all tree models, but not for DNN. Similarly, the tree-based models have the leading predictor of CO_2 storage resources as recoverable gas reservoirs[8]. The importance of the predictor variables seems to be equalized for the DNN model. Perhaps because of the complexity of the DNN algorithm that was applied, the algorithm may have detected the added influence that reservoir temperature and pressure have on the supercritical CO_2 density. For the gas reservoirs, the tree-based models show the same predictor importance ordering; recoverable gas is first, and reservoir temperature is second. Again, the gas reservoir DNN model seems to detect the separate importance that reservoir temperature and pressure have on the density of supercritical CO_2.

7 The cross-validation performance statistics for the DNN oil and gas models are presented in Supplementary Appendix 2, Table 4.

8 For the tree-based H2O models (i.e., RF and GBT), variable importances are calculated from gains in the squared error loss function over all trees. XGBoost uses a gradient in conjunction with a Hessian (Chen and Guestrin, 2016) formulation. According to Candel and LeDell (2023), the H2O DNN algorithms use the method developed by Gedeon (1997) based on weight connecting neurons in the first two hidden layers.

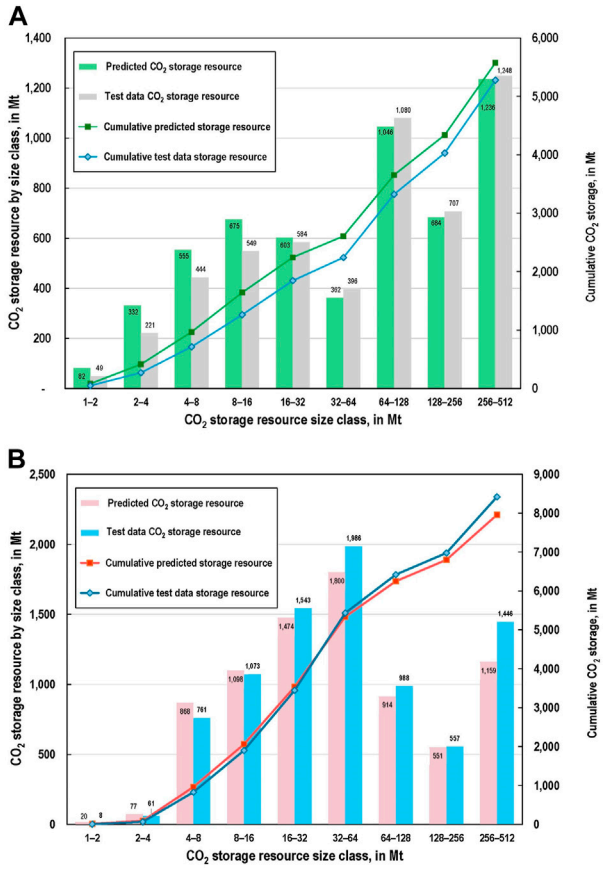

FIGURE 7
Predicted *versus* the calculated (see Supplementary Appendix 1) CO_2 storage resources assigned to: **(A)**, the 285 oil reservoirs of the test data set, and **(B)**, the 383 gas reservoirs of the test data set. Solid line shows cumulative resource growth across storage size classes and bars shows storage volume attributed to each size class. Test data predictions used the deep neural network (DNN) models (see Table 2) trained with the training data sets for oil and gas reservoirs, respectively. The size class doubles on the x-axis. Box plots of the distributions of predictor variables and CO_2 storage variables used in the machine learning (ML) models for: **(A)**, oil reservoirs, and **(B)**, natural gas reservoirs, respectively (psia, pounds per square inch absolute; MMBO, million barrels oil; BCF, billion cubic feet; Mt, megatons or million metric tons.).

TABLE 5 Ordered relative importance (percent explanation) of predictor variables for Random Forest (RF), Gradient Boosting Trees (GBT), Extreme Boosting (XGBoost), and Deep Neural Network (DNN) algorithms for predicted reservoir CO_2 storage resource.

Predictor	RF	GBT	XGBoost	DNN
Oil reservoirs				
Recoverable oil	74.0	78.0	42.2	19.3
Associated gas	23.6	15.5	24.4	21.4
Reservoir pressure	1.4	2.1	17.3	22.9
API gravity	0.6	3.2	6.1	16.0
Reservoir temperature	0.4	1.2	10.0	20.5
Gas reservoirs				
Recoverable gas	91.3	89.5	51.0	27.7
Reservoir temperature	4.8	6.1	22.4	34.1
Reservoir pressure	2.9	1.1	15.4	19.1
Specific gravity of gas	1.1	3.3	11.2	19.1

6 Summary and extensions

Conventional reconnaissance-type estimates of CO_2 storage resources of known oil and gas reservoirs require complicated engineering calculations involving 1) estimates of recoverable oil and gas, 2) reservoir properties (depth, temperature, pressure, etc.), and 3) the physical properties of the fluids. In lieu of the engineering calculations, this study examines the predictive performance of 4 ML algorithms to by-pass these computations to yield more rapid estimates of CO_2 storage resources using the same inputs. We demonstrate this capability for oil and gas reservoirs located in western Europe. The predictive performance of the DNN algorithm was superior to those of the RF, GBT, and XGBoost algorithms as measured by the root mean square errors, explained variation of the target variable, and mean absolute error statistics. The root mean square error estimates for all models (in Table 4) appear somewhat high relative to the mean CO_2 resource estimates shown in Table 1, 2. This is likely induced by the skewed size-frequency distributions of the oil and gas reservoirs and their estimates of CO_2 storage resources. Future analysis may focus on ways to partition the data to reduce these errors in predicting extreme values. In particular, there may be value in developing separate models for onshore and offshore reservoirs due to differences in physical characteristics, dispositional environments, and related operational and management factors.

Whereas reconnaissance estimates of CO_2 storage resources may be useful to answer broad planning questions they are not sufficiently sophisticated for site specific evaluation (Wen et al., 2023). Such evaluations involve much more detailed spatial data to be used for reservoir modeling (Erten et al., 2023). Future studies could lead to the potential combining of geostatistical approaches to ML. Da Silva (2020) provides a comparative analysis of ML and geostatistics for mineral resource estimation.

The ML methods and models developed and presented here may be important to the energy transition and the climate change debate because they provide a more rapid and direct path for a reconnaissance assessment of subsurface carbon storage resources. The work presented here represents an effective means for operators, researchers, and policymakers alike to rapidly and accurately quantify the potential storage resource.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Part of the data are available for purchase from IHS Markit (now part of S&P Global). Part is available from the National Institute of Standards. Requests to access these

datasets should be directed to https://energyportal.ci.spglobal.com/home and https://webbook.nist.gov/chemistry/fluid.

Author contributions

EA: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review and editing. PF: Data curation, Resources, Software, Visualization, Writing – review and editing, Funding acquisition. TC: Methodology, Supervision, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. For Attanasi and Freeman, this work was funded by the U.S. Geological Survey's Energy Resources Program. Research and preparation of paper were done as part of official duties assigned by the U.S. Geological Survey, United States Government. Coburn was supported by the Department of Systems Engineering, Colorado State University.

Acknowledgments

Authors gratefully acknowledge reviews of the manuscript by Marc Buursink, Peter Warwick, and Clinton Scott, all of the U.S. Geological Survey. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Conflict of interest

The 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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2025.1562087/full#supplementary-material

References

Abdulwarith, A., Ammar, M., and Dindoruk, B. (2024). "Prediction/assessment of CO_2 EOR and storage efficiency in residual oil zones using machine learning techniques," in *Carbon capture, utilization, and storage conference (CCUS), Houston (march 11-13), CCUS:4011705*. Available online at: https://ccusevent.org/portals/32/abstracts/4011705.pdf.

Ajayi, T., Gomes, J. S., and Bera, A. (2019). A review of CO_2 storage in geological formations emphasizing modeling, monitoring and capacity estimation approaches. *Petroleum Sci.* 16, 1028–1063. doi:10.1007/s12182-019-0340-8

Aminu, M. D., Nabavi, S. A., Rochelle, C. A., and Manovic, V. (2017). A review of developments in carbon dioxide storage. *Appl. Energy* 208, 1389–1419. doi:10.1016/j.apenergy.2017.09.015

Anthonsen, K. L., and Christensen, N. P. (2021). EU geological CO_2 storage summary. Prepared by the geological Survey of Denmark and Greenland for clean air task force (revised, oct 2021). Danmarks og grønlands geologiske undersøgelse rapport; no. 34. *GEUS*. doi:10.22008/gpub/34594

Argatan, E., Gaddipati, M., Yip, Y., Savage, B., and Ozgen, C. (2018). CO_2 storage in depleted oil and ga fields in the Gulf of Mexico. *Int. J. Greenh. Gas Control* 72, 38–48. doi:10.1016/j.ijggc.2018.02.022

Bachu, S. (2008). Comparison between methodologies recommended for estimation of CO2 storage capacity in geological media. Phase III report, carbon sequestration leadership forum (CSLF) task force on CO2 storage capacity estimation and USDOE capacity and fairways subgroup, regional carbon sequestration partnerships program, CSLF-T-2008-4. Available online at: https://fossil.energy.gov/archives/cslf/sites/default/files/documents/PhaseIIIReportStorageCapacityEstimationTaskForce0408.pdf.

Bachu, S., Bonijoly, D., Bradshaw, J., Burruss, R., Holloway, S., Christensen, N. P., et al. (2007). CO_2 storage capacity estimation: methodology and gaps. *Int. J. Greenh. Gas Control* 9, 430–443. doi:10.1016/S1750-5836(07)00086-2

Bakhshian, S., Shariat, A., and Raza, A. (2022). Assessment of CO_2 storage potential in reservoir with residual gas using deep learning. *Interpretation* 10 (3), SG37–SG46. doi:10.1190/int-2021-0147.1

Bradshaw, J., Bachu, S., Bonijoly, D., Burruss, R., Holloway, S., Christensen, N. P., et al. (2006). CO_2 storage capacity estimation: issues and development of standards. *Int. J. Greenh. Gases Control* 1, 62–68. doi:10.1016/s1750-5836(07)00027-8

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and regression trees*. Boca Raton, FL: Chapman & Hall, 358. Available online at: https://books.google.com/books/about/Classification_and_Regression_Trees.html?id=JwQx-WOmSyQC.

Candel, A., and LeDell, E. (2023). Deep learning with H2O, 6th ed. H2O. Available online at: https://www.h2o.ai/resources/booklet/deep-learning-with-h2o.

Cantucci, B., Buttinelli, M., Procesi, M., Sciarra, A., and Anselmi, M. (2016). "Algorithms for CO_2 storage capacity estimation: review and case study," in *Geologic carbon sequestration*. Editors V. Vishal, and T. Singh (Cham: Springer). doi:10.1007/978-3-319-27019-7_2

Carolus, M., Biglarbigi, K., Warwick, P. D., Attanasi, E. D., Freeman, P. A., and Lohr, C. D. (2018). *Overview of a comprehensive resource database for the assessment of recoverable hydrocarbons produced by carbon dioxide enhanced oil recovery*. Reston, VA: U.S. Geological Survey Techniques and Methods, 31. doi:10.3133/tm7C16

Charpentier, R. R., and Klett, T. R. (2005). Guiding principles of USGS methodology for assessment of undiscovered conventional oil and gas resources. *Nat. Resour. Res.* 14, 175–186. doi:10.1007/s11053-005-8075-1

Chen, B., Harp, D. R., Lin, Y., Keating, E., and Pawar, R. (2018). "Application of machine learning technique in CO_2 storage and enhanced oil recovery," in *Poster presentation, IMA workshop on recent advances in machine learning and computational*

methods for geoscience. Los Alamos National Laboratory. Available online at: https://www.researchgate.net/publication/328543040_Application_of_Machine_Learning_Techniques_in_CO2_Storage_and_Enhanced_Oil_Recovery.

Chen, B., and Pawar, R. (2019). Capacity assessment and co-optimization of CO_2 storage and enhanced oil recovery in residual oil zones. *J. Petroleum Sci. Eng.* 182, 106342. doi:10.1016/j.petrol.2019.106342

Chen, T., and Guestrin, C. (2016). "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (New York: KDD '16), 785–794. doi:10.1145/2939672.2939785

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al. (2019). Extreme gradient boosting: package 'xgboost. *R. package*. Available online at: https://CRAN.R-project.org/package=xgboost.

Consoli, C. (2016). *Global storage portfolio: a global assessment of the geological storage resource potential*. Melbourne, VIC: Global CCS Institute. Available online at: https://www.globalccsinstitute.com/resources/publications-reports-research/global-storage-portfolio-a-global-assessment-of-the-geological-co2-storage-resource-potential.

da Silva, R. A. (2020). A comparative analysis between the geostatistics and machine learning methods for mineral resource estimation. *Int. J. Geoscience, Eng. Technol.* 2, 14–22. doi:10.70597/ijget.v2i1.380

Dumakor-Dupey, N. K., and Arya, S. (2021). Machine learning—a review of applications in mineral resource estimation. *Energies* 14, 4079. doi:10.3390/en14144079

Erten, G. E., Erten, O., Karacan, C. O., Boisvert, J., and Deutsch, C. V. (2023). Merging machine learning and geostatistical approaches for spatial modeling of geoenergy resources. *Int. J. Coal Geol.* 276, 104328. doi:10.1016/j.coal.2023.104328

Ferreira, C. A. S., Kadeethum, T., Amour, F., Hosseindeh, B., Abdollahi, A., Calvert, A. S., et al. (2024). "A deep learning model for CO_2 storage in a depleted ga reservoir using parse well data," in *Proceedings, 85th EAE annual conference and exhibition*, 1–5. Available online at: https://www.earthdoc.org/content/papers/10.3997/2214-4609.2024101572.

Friedman, J. H. (2002). Stochastic gradient boosting. *Comput. Statistics Data Analysis* 38, 367–378. doi:10.1016/s0167-9473(01)00065-2

Gedeon, T. D. (1997). Data mining of inputs: analysing magnitude and functional measures. *Int. J. Neural Syst.* 8, 209–218. doi:10.1142/s0129065797000227

Goodman, A. (2012). "Comparison of CO_2 storage resource methodologies," in *Oral presentation, carbon storage R&D Project review meeting, developing the technologies and building the infrastructure for CCUS*. Pittsburgh, PA: National Energy Technology Laboratory.

Goodman, A., Hakala, A., Bromhal, G., Deel, D., Rodosta, T., Frailey, S., et al. (2011). U.S. DOE methodology for the development of geologic storage potential for carbon dioxide at the national and regional scale. *Int. J. Greenh. Gas Control* 5, 952–965. doi:10.1016/j.ijggc.2011.03.010

Goodman, A., Sanguinito, S., and Levin, J. S. (2016). Prospective CO_2 saline resource estimation methodology: refinement of existing US-DOE-NETL methods based on data availability. *Int. J. Greenh. Gas Control* 54, 242–249. doi:10.1016/j.ijggc.2016.09.009

Gorecki, C. D., Ayash, S. C., Liu, G., Braunberger, J. R., and Dotzenrod, N. W. (2015). A comparison of volumetric and dynamic CO_2 storage resource and efficiency in deep saline formations. *Int. J. Greenh. Gas Control* 42, 213–225. doi:10.1016/j.ijggc.2015.07.018

H2O.ai (2022). H2O, Open-source machine learning platform for the enterprise. Available online at: https://h2o.ai/platform/ai-cloud/make/h2o (Accessed January, 2023).

Haagsma, A., Main, J., Pasumarti, A., Valluri, M., Scharenberg, M., Larsen, G., et al. (2020). A comparison of carbon dioxide storage resource estimate methodologies for a regional assessment of the Northern Niagaran Pinnacle Reef Trend in the Michigan Basin. *Environ. Geosci.* 27, 9–23. doi:10.1306/eg.11051919019

Halder, N. (2023). Harnessing the power of grid search for optimized machine learning models. *Medium*. Available online at: https://medium.com/analysts-corner/harnessing-the-power-of-grid-search-for-optimized-machine-learning-models-5878e3abf2d6.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. Springer Series in Statistics, Springer.

He, X., Zhu, W., AlSinan, M., Kwak, H., and Hoteit, H. (2022). *CO_2 storage capacity prediction in deep saline aquifers: uncertainty and global sensitivity analysis*. Riyadh (February: International Petroleum Technology Conference. Available online at: https://onepetro.org/IPTCONF/proceedings-abstract/22IPTC/2-22IPTC/D021S044R003/480031.

Heidug, W. (2013). *Methods to assess geologic CO_2 storage capacity: status and best practice*. Paris, France: International Energy Agency. Available online at: https://www.osti.gov/etdeweb/servlets/purl/22134453.

IHS Markit (2021). *International petroleum exploration and production database*. London: IHS Markit.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). *An introduction to statistical learning*. 2nd ed. Available online at: https://hastie.su.domains/ISLR2/ISLRv2_corrected_June_2023 (Accessed August, 2022).

Jones, M. M., Wiens, A., Buursink, M., Brennan, S., Freeman, P., Varela, B., et al. (2024). "A methodology to estimate CO_2 and energy gas storage resources in depleted conventional gas reservoirs," in *17th greenhouse gas control technologies conference* (Calgary, Canada: GHGT-17) 20th-24th October 2024). Available online at: https://ssrn.com/abstract=5014690.

Kadeethum, T., Verzi, S. J., and Yoon, H. (2023). *Efficient machine-learning surrogates for large-scale geological carbon and energy storage*. Sandia National Laboratories. doi:10.48550/arXiv.2310.07461

Kearns, J., Teletzke, G., Palmer, J., Thomann, H., Kheshgi, H., Chen, Y.-H. H., et al. (2017). Developing a consistent database for regional geologic CO_2 storage capacity worldwide. *Energy Procedia* 114, 4697–4709. doi:10.1016/j.egypro.2017.03.1603

Lyons, W. C. (1999). *Standard handbook of petroleum and natural gas engineering*. Houston, Texas: Gulf Publishing Company.

Mahoob, M. A., Celik, T., and Genc, B. (2022). Review of machine learning-based mineral resource estimation. *J. South. Afr. Inst. Min. Metallurgy* 122, 655–664. doi:10.17159/2411-9717/1250/2022

Moore, R. (2022). "CO_2 resources and their development," in *International energy agency, energy Technology policy division* (Carbon Capture, Utilization and Storage Unit). Paris, France: International Energy Agency. Available online at: https://www.iea.org/reports/co2-storage-resources-and-their-development.

National Energy Technology Laboratory (NETL) (2015). "CO_2 storage resource methodology," in *Carbon storage atlas*. 5th ed. Available online at: https://netl.doe.gov/node/5964.

National Institute of Standards and Technology (NIST) (2024). NIST Chemistry, SRD 69, Thermophysical properties of fluid systems. Available online at: https://webbook.nist.gov/chemistry/fluid/(Accessed September, 2024).

Ouyang, L. (2011). New correlations for predicting the density and viscosity of supercritical carbon dioxide under conditions expected in carbon capture and sequestration operations. *Open Petroleum Eng. J.* 4, 13–21. doi:10.2174/1874834101104010013

Peck, W. D., Azzolina, N. A., Ge, J., Gorecki, C. D., Gorz, A. J., and Melzer, S. (2017). Best practices for quantifying the CO_2 storage resource estimates in CO2 enhanced oil recovery. *Energy Procedia* 114, 4741–4749. doi:10.1016/j.egypro.2017.03.1613

Pingping, S., Xinwei, L., and Qiujie, L. (2009). Methodology for estimation of CO_2 storage capacity in reservoirs. *Petroleum Explor. Dev.* 36, 216–220. doi:10.1016/s1876-3804(09)60121-x

Popova, O., Small, M., Thomas, A. C., McCoy, S. T., Karimi, B., Goodman, A., et al. (2012). Comparative analysis of carbon dioxide storage resource assessment methodologies. *Environ. Geosci.* 19, 105–124. doi:10.1306/eg.06011212002

Sanguinito, S., Singh, H., Myshakin, E. M., Goodman, A. L., Dilmore, R. M., Grant, T. C., et al. (2020). Methodology for estimating the prospective CO_2 storage resource of residual oil zones at the national and regional scale. *Int. J. Greenh. Gas Control* 96, 103006. doi:10.1016/j.ijggc.2020.103006

SAS Institute (2025). Machine learning: what it is and why it matters. Available online at: https://www.sas.com/en_us/insights/analytics/machine-learning.html.

Smeenk, S., and Leeuwenburgh, O. (2023). Initial evaluation of a machine learning proxy for CO_2 -storage in depleted gas fields. *Proc. Fifth EAGE Conf. Petroleum Geostatistics*, 1–5. doi:10.3997/2214-4609.202335057

Society of Petroleum Engineers (2017). CO2 storage resources management system, v 1.02, 50. Available online at: https://www.spe.org/en/industry/co2-storage-resources-management-system/.

Span, R., and Wagner, W. (1999). A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100K at pressures up to 800 Mpa. *J. Phys. Chem. Reference Data* 25, 1509–1596. doi:10.1063/1.555991

Standing, M. B. (1948). "A pressure-volume-temperature correlation for mixtures of California oils and gases," in *Drilling and production practice, 1947* (New York: American Petroleum Institute and Society of Petroleum Engineers), 275–287. Available online at: https://www.onepetro.org/conference-paper/API-47-275 (Accessed May 11, 2015).

Standing, M. B., and Katz, D. L. (1942). *Density of natural gases: transactions of the American Institute of mining Engineers (AIME), society of petroleum engineer. SPE–942140-G*, 10.

Thanh, H. V., Sugai, Y., and Sasaki, K. (2020). Application of artificial neural network for predicting the performance of CO_2 enhanced oil recovery and storage in residual oil zones. *Sci. Rep.* 10, 18204. doi:10.1038/s41598-020-73931-2

U.S. Energy Information Administration (EIA) (2000). *U.S. crude oil, natural gas, and natural gas liquids reserves—1999 annual report*. Washington, DC: U.S. Energy Information Administration, DOE/EIA–021699. Available online at: https://www.eia.gov/naturalgas/crudeoilreserves/archive/1999/full.pdf.

U.S. Geological Survey (USGS) (2013). National assessment of geologic carbon dioxide storage resources—summary. Fact Sheet 2013-3020, Version 1.1. Available online at: https://pubs.usgs.gov/fs/2013/3020/pdf/fs2013-3020_508.pdf.

Wen, G., Li, Z., Long, Q., Azizzadenesheli, K., Anandkumar, A., and Benson, S. M. (2023). Real-time high-resolution CO_2 geological storage prediction using nested Fourier neural operators. *Energy and Environ. Sci.* 16, 1732–1741. doi:10.1039/D2EE04204E

Wichert, E., and Aziz, K. (1971). Compressibility factor of sour natural gases. *Can. J. Chem. Eng.* 49, 267–273. doi:10.1002/cjce.5450490216

Wilpon, J., Thomson, D., Bangalore, S., Haffner, P., and Johnston, M. (2017). The fundamentals of machine learning. *Whitepaper, Interact.* Available online at: https://www.interactions.com/resources/technology/fundamentals-machine-learning.

Xing, W., and Du, D. (2019). Dropout prediction in MOOCs: using deep learning for personalized intervention. *J. Educ. Comput. Res.* 57, 547–570. doi:10.1177/0735633118757015

You, J., Ampomah, W., Kuatsienyo, E. J., Sun, Q., Balch, R. S., Aggrey, W. N., et al. (2019). "Assessment of enhanced oil recovery and CO2 storage capacity using machine learning and optimization framework," in *Proceedings, 81st EAGE conference and exhibition* (London: SPE-195490-MS). Available online at: https://onepetro.org/SPEEURO/proceedings-abstract/19EURO/4-19EURO/217905.

Zhang, Y., Jackson, C., and Krevor, S. (2022). An estimate of the amount of geological CO_2 storage over the period of 1996-2020. *Environ. Sci. Technol. Lett.* 9, 693–698. doi:10.1021/acs.estlett.2c00296

Zhao, X., Liao, X., Wang, W., Chen, C., Rui, Z., and Wang, H. (2014). The CO_2 storage capacity evaluation: methodology and determination of key factors. *J. Energy Inst.* 87, 297–305. doi:10.1016/j.joei.2014.03.032