Journal Homepage: http://ijcpe.uobaghdad.edu.iq
Iraqi Journal of Chemical and Petroleum Engineering
Vol. 27 No. 2 (June 2026) 19 – 32
EISSN: 2618-0707, PISSN: 1997-4884

AI-driven carbon-neutral gas recovery through CO_2 injection and storage integration

Mohamedgifon Eissa [a, *], Manash Protim Mudoi [b]

a Department of Petroleum Engineering, University of Petroleum and Energy Studies (UPES), Dehradun, 248001, India
b Department of Chemical Engineering, University of Petroleum and Energy Studies (UPES), Dehradun, 248001, India

Abstract

The rising level of Carbon Dioxide (CO_2) concentration and the growing energy demand of the world have given a desperate necessity of carbon-neutral and sustainable energy sources. Carbon Capture, Utilization and Storage (CCUS), especially CO_2-Enhanced Gas Recovery (CO_2-EGR) has become an attractive technology to maximize hydrocarbon recovery with a minimal environmental impact. The paper introduces an AI-based system of recovering carbon-neutral gas by implementing CO_2 gas injection and storage. A number of supervised learning models, including XGBoost, Random Forest (RF), and a hybrid RF-XGBoost, were used in conjunction with data preprocessing and feature engineering to come up with the results from the NETL CCS dataset. They also use Bayesian Optimization to adjust the parameters. The findings reveal that the hybrid model does better than the individual models because it has the lowest MAE (0.080), MSE (0.011), and RMSE (0.105), as well as the highest value of R^2 (0.96). Further, the recovery efficiency increases by 83 to 92 percent, the leakage risk is lowered by 0.14 to 0.06, and the cost of operation is lowered by 1.00 to 0.72. The results indicate that AI-based strategies are effective to maximize the efficiency, safety, and sustainability with a carbon-neutral gas recovery system.

Keywords: Carbon Capture and Storage (CCS); CO_2-Enhanced Gas Recovery (CO_2-EGR); Machine Learning; XGBoost; Bayesian Optimization.

Received on 03/02/2026, Received in Revised Form on 12/05/2026, Accepted on 12/05/2026, Published on 30/06/2026

https://doi.org/10.31699/IJCPE.2026.2.2

1- Introduction

The increasing effects of climate change, along with the ever-increasing energy demand, have increased the need for a sustainable energy solution [1]. Natural gas and other fossil fuels continue to play an important part in satisfying energy demand because of their accessibility, existing infrastructure, and reduced CO_2 emissions, even if the usage of renewable energy is on the rise. Nevertheless, the use of natural gas generates a huge amount of CO_2, which is responsible for increasing global warming and damaging the environment [2]. Thus, there is a need to develop effective strategies to balance energy and environmental sustainability [3, 4].

In light of global net-zero ambitions and international climate agreements like the Paris Agreement, CCUS has become an important technical strategy for cutting GHG emissions [5]. Among CCUS approaches, CO_2-EGR has gained considerable attention. This process involves injecting the captured CO_2 into depleted or partially depleted gas reservoirs, thus enhancing hydrocarbon recovery, as well as providing long-term geological storage of CO_2 [6]. Therefore, this technique provides a double advantage in terms of efficiency and carbon emission reduction. The integration of energy production and carbon storage makes this technique a potential candidate for carbon-neutral gas recovery. The entire process of CO_2 capture, transportation, injection, enhanced gas recovery, and geological storage is shown in Fig. 1.

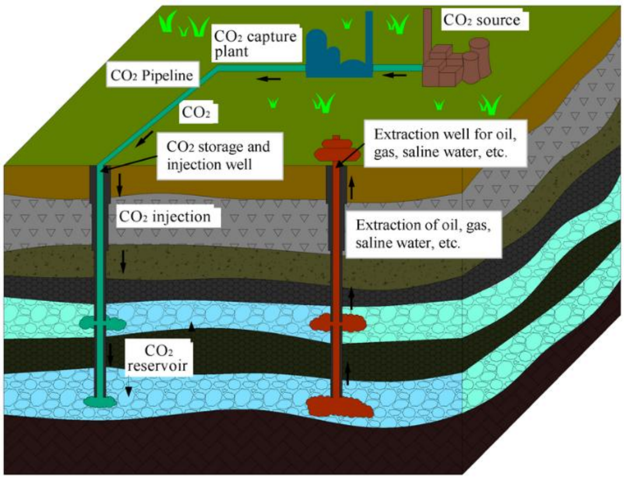

Fig. 1. Schematic representation of CO_2 capture, transportation, injection, enhanced gas recovery (EGR), and geological storage in subsurface reservoirs [7]

The process of capturing and storing CO_2 underground results in operational and technical difficulties despite its

*Corresponding Author: Email: mohamedgifon4@gmail.com

potential advantages [8]. The complexity of reservoir systems arises from their various rock characteristics and the unpredictable behavior of fluids and the unknown aspects of their geological structure. The traditional approach to reservoir simulation achieves accurate results but needs extensive computational resources and struggles with handling big data and making instant choices [9, 10]. The dynamic behavior of subsurface processes creates difficulties for precise forecasting of critical performance metrics which include recovery efficiency and CO_2 leakage probability [11].

Artificial Intelligence (AI) and Machine Learning (ML) have become the effective tools of preventing the existing issues of the field of energy systems. The AI-powered models leverage their capabilities to handle large and diverse datasets as they find complex patterns and develop nonlinear relationships that current approaches struggle to model [12]. The application of AI to CO_2-EGR involves making the future value of a reservoir workable and perfecting the technique of injecting the oil and assessing the safety of a storage facility. Deep learning (DL) combined with the methods of ensemble learning and reinforcement learning produce systems that learn by analyzing the data to make decisions that increase the performance of operating schedules and safer environmental activities [13, 14]. The AI-based models rely on various types of data including geological data, well log data and pressure and temperature measurements and production recordings to generate accurate and reliable forecasts [15]. The real-time monitoring is supported by the models as they are capable of detecting anomalies in sensor data that assist in identifying the risk of leakage and system failure. AI enables businesses to reduce expenses by managing resources more effectively and using energy more efficiently during injection processes and reduced chances of operational issues [16].

However, the increasing use of AI in CCUS and reservoir engineering research is characterized by multiple research areas that need to be explored as discussed in reference [17]. The research is characterized by the failure to display transparent data representation that leads to undefined input and output relationships. The research is also characterized by the failure to make comparisons between the new modeling approaches and existing ML models such as RF and XGBoost. The models become less reliable and unsuitable for practical applications as researchers fail to focus on two important issues that include uncertainty quantification and validation of models using real data.

This paper provides an AI-based framework for carbon-neutral gas recovery that integrates CO_2 injection and storage to solve these concerns. In this framework, there is a focus on the prediction and optimization of the parameters associated with gas recovery efficiency. Moreover, a comparison with existing ML benchmarks is carried out to assess the efficiency of the introduced framework. Improving decision-making and efficiency and creating energy systems that do not produce any carbon dioxide are the goals of the framework's development. The following are the research objectives as shown below:

- To develop an AI-driven framework for carbon-neutral gas recovery through the integration of CO_2 injection and storage processes.
- To analyze and preprocess CCS project data from the NETL database for effective machine learning-based modeling.
- To implement and compare ML models, including RF, XGBoost, and a hybrid RF–XGBoost model, for predicting recovery efficiency and leakage risk.
- To optimize operational parameters such as injection rate, pressure, and cost using Bayesian Optimization for improved system performance.

To evaluate the effectiveness of the proposed framework using performance metrics (MAE, MSE, RMSE, and R^2) and validate its capability in enhancing recovery efficiency while minimizing environmental risk and operational cost.

2- Review of literature

In recent years, AI has introduced new technologies that allow making predictions more accurately, as well as more effective ways of storing CO_2 in underground deposits that scientists believe are crucial to the realization of the carbon-neutral natural gas production system. There has been an increasing trend of using ML techniques to develop models that mimic complex geological structures and fluid interactions that are too computationally intensive to be simulated using the conventional simulation techniques. RF and XGBoost tree-based models have demonstrated effective performance since Wen et al. [18] utilized them to predict CO_2 storage capacity based on reservoir characteristics such as porosity and permeability and pressure and temperature. Yan et al. [19] demonstrated that DL systems using neural networks have the ability to simulate the flow of CO_2 through porous materials effectively due to the ability of the system to accurately predict the saturation patterns in different injection conditions. Yang et al. [20] established the fact that physics-informed neural networks have enhanced prediction accuracy since the physical laws that govern the flow of fluids are incorporated into the process of learning. Abdellatif et al. [21] used the high-resolution pore-scale datasets to come up with AI models that forecast CO_2 water interactions in different geological formations. The Convolutional Neural Networks (CNNs) developed by Zhu et al. [22] are used to analyze the CO_2 plume movement patterns on a spatial scale that offer superior visualization and forecasting of their subsurface behaviour. Iddrisu et al. [23] applied gradient boosting models to develop forecasts regarding the operation of long-term CO_2 trapping mechanisms that encompasses residual and solubility trapping strategies that safeguard carbon storage sites. The research indicates that AI has the capability to handle various tasks since it demonstrates the possibility of addressing complex problems that demand sophisticated thought processes.

Researchers have applied AI-based techniques to enhance gas extraction results through CO_2 injection into unconventional rock formations which include shales and coal seams. The researchers Amirkhani et al. [24] applied supervised learning models which include artificial neural networks and regression-based methods to predict methane recovery from shale gas reservoirs that use CO_2 injection. The researchers Gu et al. [25] combined Support Vector Machines (SVM) and Multi-Layer Perceptron (MLP) models with reservoir simulation data to predict coal seam gas production, achieving accurate results while identifying key factors which included permeability, pressure, and injection rate. The researchers Zentou et al. [26] used ensemble learning methods to improve CO_2 injection methods which achieve maximum gas recovery while reducing operational expenses. The researchers Davila et al. [27] utilized machine learning models to forecast oil production and CO_2 storage efficiency in Water-Alternating-Gas (WAG) injection systems which operate with better speed and similar precision to existing numerical simulation methods. The researchers used data-driven methods to study how heterogeneous reservoirs affect gas displacement efficiency, proving that AI models can represent complicated nonlinear patterns present in underground systems which demonstrates that AI technologies help improve the operational performance and financial sustainability of CO_2-based gas extraction systems.

More recent research has focused on the use of AI in real time optimization, monitoring, and assurance of safety in CO_2 injection and storage systems in addition to prediction and recovery. Zhou et al. [28] designed the reinforcement learning techniques to dynamically optimize the injection strategies by relaxation of the operation parameters of pressure and flow rates based on the dynamics of the reservoir condition. Time-series models have been used to forecast reservoir pressure and CO_2 plume formation and act proactively and mitigate risks, including the Long Short-Term Memory (LSTM) networks. Mashhadimoslem et al. [29] have also been applied to detect channels of CO_2 leaking in seismic data using DL techniques such as CNNs, which is more precise, and can, therefore, enhance the safety and integrity of the storage systems. Hybrid AI systems integrating ML and conventional simulation models have also been suggested in order to optimize simultaneously the gas recovery and the carbon storage, to reach the balance between both economic performance and environmental sustainability. In general, the combination of AI with the CO_2 injection and storage technologies offers a strong and prospective solution to facilitate the carbon-neutral gas recovery that can overcome the issues associated with the efficiency, safety, and the reliability of the long-term storage.

3- Research methodology

The overall research methodology for AI-based carbon-neutral gas recovery using injected CO_2 in the gas reservoir is shown in Fig. 2. It starts with the dataset provided by NETL CCS, followed by data preprocessing, feature engineering, ML modeling involving RF, XGBoost, and a hybrid approach. The model's parameters are fine-tuned using Bayesian Optimization, and its performance is evaluated using an array of measures.

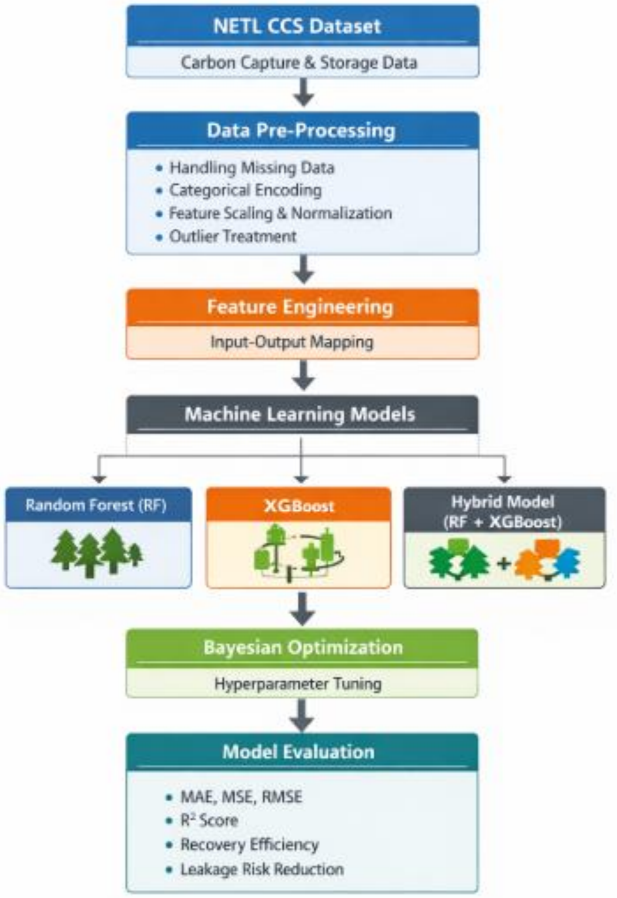

Fig. 2. Proposed methodology

3.1. Dataset used

An extensive and well-organized database on CCS projects worldwide, the National Energy Technology Laboratory (NETL) CCS Database is the source of the data included in this research. This publicly available database is a compilation of data on industrial, governmental, and research-based sources providing information on all ongoing, planned, and completed CCS projects around the globe. It contains such important characteristics as project description, country, type of a facility, capture technology, its operational status, estimated project cost, and the volume of CO_2 captured and stored. Moreover, the database includes data on the evaluation of storage sites and the strategies of their deployment in the context of CO_2 sequestration. The latest records show that it has more than 300 CCS projects spread across over 30 countries and six continents, including capture, storage and integrated operations. This heterogenous and structurally arranged data is an effective source of analyzing the trends in CCS deployment and AI-based modeling of carbon-neutral power systems. The final dataset used to train and assess the model has 305 samples and 6 input characteristics. The dataset includes various CCS projects worldwide, and a sample of representative entries is shown in Table 1.

Table 1. Sample entries from the NETL CCS dataset

Project Name	Country	Facility Type	Capture Technology	Capacity ($MtCO_2$/yr)	Project Status
Boundary Dam CCS	Canada	Power Plant	Post-combustion	1.0	Operational
Petra Nova	USA	Power Plant	Post-combustion	1.4	Operational
Sleipner CO_2 Storage	Norway	Gas Processing	Pre-combustion	0.9	Operational
Gorgon CO_2 Project	Australia	LNG Processing	Pre-combustion	3.4	Operational
Illinois Industrial CCS	USA	Ethanol Plant	Pre-combustion	1.0	Active

3.2. Data pre-processing

Preprocessing the data is an important step to ensure its quality, consistency, and reliability before applying ML models. The NETL CCS dataset consists of project-level metadata collected from different sources, and it may contain missing values as well as variations in scale.

* Handling missing data

There are also missing values in the dataset which are managed with proper imputation methods to create complete and consistent data. Numerical features are mean imputed whereas mode imputed in categorical variables. The strategy could be used to preserve the data's wide distribution preserved and to avoid losing vital information.

Mean imputation for numerical features is defined as:

$$X_i = \frac{1}{n}\sum_{k=1}^{n} X_k \tag{1}$$

where n is the count of accessible observations and X_i is the value that is imputed.

For categorical variables, mode imputation is applied as:

$$X_{\text{mode}} = \arg\max(f(X)) \tag{2}$$

where $f(X)$ denotes the frequency of occurrence of each category.

Additionally, nominal variables such as facility type and technology for capture are converted into numerical form using one-hot encoding to be able to use them in ML models.

* Categorical encoding

The ML model requires numerical data as input. Thus, the categorical variables, facility types, and capture technology are subjected to one-hot encoding. A categorical variable with m categories can be represented as:

$$X = [x_1, x_2, \dots, x_m], x_i \in \{0,1\} \tag{3}$$

where only one element is 1 and the rest are 0.

* Feature scaling and normalization

Since the dataset contains features with varying magnitudes (e.g., capacity, cost, pressure), normalization is applied to ensure uniform scaling. The Min-Max normalization technique is used:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4}$$

The numerical stability and convergence speed of the trained model are both enhanced by this transformation, which scales all characteristics to the interval [0,1].

* Outlier treatment and distribution adjustment

Outliers can distort model predictions and reduce generalization performance. To address this, statistical normalization is applied using standardization:

$$Z = \frac{X - \mu}{\sigma} \tag{5}$$

Here is where the feature's standard deviation (σ) and mean (μ) are defined. The data could be transformed to ensure a normal distribution with a mean of 0 and a variance of 1.

* Data consistency and feature alignment

All features are aligned and validated to ensure they have the same units and formats. This step holds vital importance for CCS datasets because their attributes originate from multiple industrial sectors and different geographical locations.

3.3. Data splitting

To determine the level of performance of their model, the researchers designed two different sets of data to be used in training and testing. Typically, training uses around 70% of the whole data, while testing makes use of the remaining 30%. The training data also has a section that serves as validation set that is used to optimize model parameters without overfitting. The data division is structured in such a way that it facilitates the process of effective model training that leads to superior performance on unheard data.

3.4. Feature engineering and input–output mapping

The concept of feature engineering is a significant process in ML since it transforms raw data into useful and informative forms thereby improving prediction and generalization of ML models. In the context of the CCS systems, feature engineering can be used to find the patterns of interest in the project-level data and to model the intricate relationships between the operation parameters and the indicators of performance, such as the efficiency of the recovery and the leakage risk.

The input variables are constructed, basing on the critical characteristics acquired in the NETL CCS database, including the CO_2 capture capacity, the kind of the facility, the technology deployed to extract the capture, the project status, and the geographical location. These characteristics represent quantitative and qualitative characteristics of CCS projects. Categorical variables are converted into numbers in such a way that they can be fit in machine learning models, and the numerical attribute is normalized during the preprocessing stage. The resulting data set has a systematic description of CCS systems in different circumstances of functioning. Mathematically, the dataset is represented as a feature matrix:

$$X = [x_{ij}] \in \mathbb{R}^{n \times f} \tag{6}$$

where n denotes the number of CCS projects and f represents the number of features. Each row vector $x_i \in \mathbb{R}^f$ corresponds to an individual project, capturing its operational and technological attributes.

The NETL dataset does not provide direct reservoir-level performance metrics; therefore, target variables were constructed for predictive modeling. Since the NETL CCS dataset primarily contains project-level metadata and lacks dynamic reservoir measurements, the target variables, namely recovery efficiency and leakage risk, are derived using data-driven estimation techniques. These outputs are generated based on normalized relationships between key attributes such as capture capacity, facility type, operational status, and capture technology, following methodologies reported in recent CCS and machine learning studies. The study focuses on two primary continuous outputs: recovery efficiency and leakage risk. It is important to note that these target variables are not directly measured from field data but are approximated using data-driven modeling and domain-informed assumptions to enable supervised learning. The output vector is defined as:

$$Y = [y_1, y_2] \tag{7}$$

where y_1 represents recovery efficiency and y_2 represents leakage risk.

The relationship between the input variables and the output variables is depicted by the nonlinear mapping:

$$Y = f(X; \theta) \tag{8}$$

Where $f(\cdot)$ is the predictive model and where is the model parameters that have been trained. This formulation allows the model to embody complex interrelations in the presence of attributes of CCS project and their influence on the performance of the system. The derived features are built to improve the models' performance by combining and transforming existing variables. They are normalized capacity indices and technology-based measures that help in projecting the fluctuations in the operational efficiency. The change of features can be formulated as:

$$X' = \phi(X) \tag{9}$$

where $\phi(\cdot)$ represents the feature transformation function.

Overall, feature engineering is an essential procedure that contributes to making a model more robust, eliminating redundancy, and enables learning nonlinear association, which is essential to the adequate prediction of carbon-neutral gas recovery performance and CO_2 storage properties.

3.5. Machine learning models and prediction framework

A tree-based model's capacity to capture non-linear patterns and its proven effectiveness on structured tabular data are the deciding factors in terms of model selection. The problem of predicting recovery efficiency and leakage risk is formulated as a supervised learning problem using structured CCS data. There are nonlinear correlations between input characteristics and goal variables, and the system is able to capture them due to the multi-model architecture. Machine learning models that work best with tabular data are used to ensure robust, accurate, and generalizable predictions for carbon-neutral gas recovery systems.

* Random Forest (RF)

A powerful and accurate ensemble learning technique, RF is able to handle complex nonlinear connections and finds widespread usage in regression and classification issues [30]. In order to make a final forecast, it builds several decision trees during training and then averages their outputs [31]. This ensemble methodology lowers overfitting and enhances the performance of generalization, which is why this ensemble approach is most appropriate to tabular data (e.g., CCS project data).

The basic concept of the RF is anchored on bootstrap aggregation (bagging), in which many subsets of data are created by selecting data samples randomly and with replacement [32]. Training is performed on each subset to obtain a separate decision tree, and randomness in feature usage is employed when constructing the trees so that each tree is different [33]. Such randomness assists in the advantage that it lowers the correlation between trees as well as improves stability of the model.

Mathematically, the prediction of the RF model is given by:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (10)$$

where $h_t(x)$ stands for the t-th tree's forecast and T is the total number of decision trees. While regression issues are solved by averaging the predictions of all trees, classification problems use majority voting.

There is a recursive process that separates the input by feature values before the decision trees in the forest are constructed. The Mean Squared Error (MSE) is one statistic that could be used to evaluate the quality of a split. A definition of it would be:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (11)$$

where y_i is the actual value and $\hat{y}_i$ is the predicted value. The objective is to minimize this error at each split to improve prediction accuracy.

Random Forest is selected due to its robustness as its main analytical tool because it processes numerical and categorical information without requiring users to establish particular data distribution models [34]. The system produces automated feature importance metrics which reveal the most critical parameters that affect both recovery efficiency and leakage risk. The system operates correctly because its design allows it to handle noisy information together with outlier data which appears in real CCS data collections. The system establishes operational parameter links to performance indicators by using nonlinear feature interaction models which identify complicated relationships between these elements. The RF model creates decision trees from dataset information to produce final predictions through their combined outcomes. Fig. 3 presents the architecture of RF.

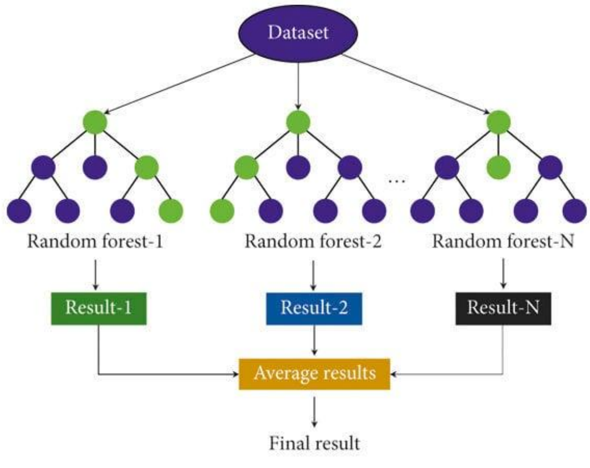

Fig. 3. RF ensemble learning process [35]

* Extreme gradient boosting (xgboost)

XGBoost is another an efficient ensemble learning technique that uses the gradient boosting approach, which has been shown to be extremely accurate in both prediction and computation [36]. XGBoost differs from other bagging-based ensemble approaches such as RF in that it employs decision trees consecutively, with each decision tree trained to rectify the errors generated by the preceding decision trees [37]. XGBoost achieves high accuracy in prediction by learning complex nonlinear relationships from the data.

The XGBoost model represents the prediction as an additive combination of multiple decision trees:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in \mathcal{F} \quad (12)$$

where $\hat{y}_i$ is the predicted output for the i-th sample, K is the number of trees, and f_k represents an individual regression tree belonging to the functional space $\mathcal{F}$.

To train the model, they minimize a regularized objective function that includes a complexity penalty and a loss function:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (13)$$

where $\Omega(f_k)$ regulates the complexity of the model and avoids overfitting, and $l(y_i, \hat{y}_i)$ is the loss function, such as the mean squared error. Penalties on the quantity of leaves and the size of leaf weights are common in regularization.

The use of gradient descent optimization is among the best strengths of XGBoost. The model computes in each iteration the loss function derivatives of the loss values and regress a new tree to the new residuals. It is the process that enables to reduce the errors successfully, along with enhancing the rate of convergence [38]. The XGBoost is particularly suitable when operating with CCS data since it is capable of assuming the heterogeneous and structured type. It is able to effectively model the correlation between the characteristics such as ability to capture, type of facility and operational parameters therefore improve the prediction on recovery efficiency and leakage risk. In addition, XGBoost has certain improved properties such as tree pruning, parallel processing and the capability to handle missing values that contribute to its performance as well. Fig. 4 shows that XGBoost model sequentially fits numerous decision trees and each decision tree offers its final prediction through correcting the previous trees.

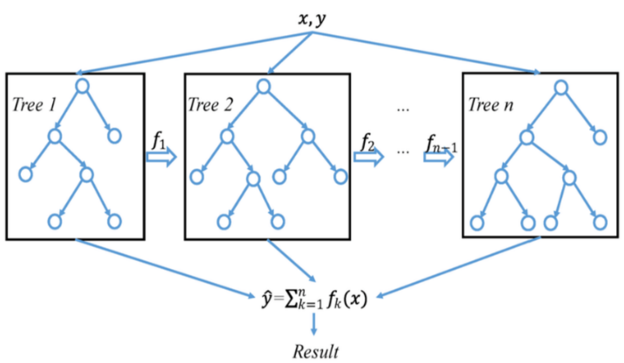

Fig. 4. XGBoost sequential tree boosting framework [39]

Another advantage of XGBoost is its capability to provide feature importance scores which permit users to determine the essential elements that affect system performance [40]. The method demonstrates its effectiveness by helping researchers assess which CCS parameters most affect carbon-neutral gas recovery.

* Hybrid Model (RF + XGBoost)

The hybrid learning is a combination of RF and XGBoost and is applied in the study to enhance accuracy in prediction and generate more powerful machine learning models. The hybridization has been applied to merge the two model systems since RF applies bagging in order to reduce variance and XGBoost applies sequential boosting in order to reduce bias. The combination of these approaches in the hybrid model enables it to have superior generalization abilities along with more reliable prediction outcomes. The framework implies the training of RF and XGBoost models individually on the same input data. The ensemble aggregation approach is used to combine the results of their predictions to produce the final output. The combination of features enables multiple feature interactions to be captured while decreasing the chances of individual model systems experiencing overfitting or underfitting.

The mathematical expression for the forecast made by the hybrid model is:

$$\hat{y} = \alpha \cdot \hat{y}_{RF} + (1 - \alpha) \cdot \hat{y}_{XGB} \quad (14)$$

where $\hat{y}_{RF}$ and $\hat{y}_{XGB}$ represent the predictions from RF and XGBoost models, respectively, and $\alpha \in [0,1]$ is a weighting parameter that controls the contribution of each model.

The value of α is empirically set based on the performance of the validation to achieve the best prediction accuracy. This weighted ensemble method enables the model to manage bias and variance effectively. For the CCS systems, the proposed hybrid model can improve the prediction of recovery efficiency and risk of leakage by considering various learning aspects. This can improve the stability of the model, avoid prediction errors, and provide a robust framework for the analysis of AI-based carbon-neutral gas recovery.

3.6. AI-driven optimization framework (bayesian optimization)

An AI-based optimization framework with the Bayesian Optimization (BO) is used to make carbon-neutral gas recovery systems more efficient and reliable. The aim of this framework is to determine the best operational parameters that maximize the recovery efficiency with the least probability of leakage and minimum cost of operation. CCS systems are complex, nonlinear and computationally expensive processes, so Bayesian Optimization can offer an effective method of global optimization under a restricted number of evaluations.

The optimization problem is modeled as a multi objective operation and the objective is to optimally optimize several performance indicators. Where x represents the decision variables (or parameters) (e.g. injection rate, pressure and operational conditions). The objective function may be formulated as:

$$\max f(x) = [f_1(x), f_2(x), -f_3(x), -f_4(x)] \quad (15)$$

where $f_1(x)$ represents recovery efficiency, $f_2(x)$ denotes CO_2 storage efficiency, $f_3(x)$ corresponds to operational cost, and $f_4(x)$ represents leakage risk. The negative signs indicate minimization objectives. For the purpose of Bayesian optimization, the objective function is approximated using a surrogate model, most often Gaussian Process Regression (GPR). The predictive mean and variance of the Gaussian Process are given by:

$$\mu(x) = k(x, X)^T [K + \sigma_n^2 I]^{-1} y \quad (16)$$

$$\sigma^2(x) = k(x, x) - k(x, X)^T [K + \sigma_n^2 I]^{-1} k(x, X) \quad (17)$$

where $k(\cdot,\cdot)$ is the kernel function, K is the covariance matrix, and σ_n^2 represents noise variance.

An acquisition function is used to direct the search for best solutions. Expected Improvement (EI) is a commonly used acquisition function that is defined as:

$$EI(x) = (\mu(x) - f(x^+))\Phi(z) + \sigma(x)\phi(z) \quad (18)$$

where $z = \frac{\mu(x) - f(x^+)}{\sigma(x)}$, $\Phi(z)$ is the cumulative distribution function, and $\phi(z)$ is the probability density function of the standard normal distribution. Here, $f(x^+)$ represents the best observed value.

Optimization is done through the iterative process of updating the surrogate model and choosing new sampling points according to the acquisition function. The method is effective in covering the search space as well as taking advantage of promising spaces. In general, the Bayesian Optimization framework facilitates effective and smart optimization of the CCS operational parameters to enhance performance of recovery, reduce risks and enable sustainable Carbon-neutral gas recovery.

3.7. Evaluation metrics

Standard regression metrics are used to evaluate the accuracy and reliability of the predictions made by the suggested ML models. Since recovery efficiency and leakage risk are continuous target variables, both error-based and statistical evaluation measures are applied. Metrics measure the model's generalizability to new data by quantifying the discrepancy between anticipated and actual values. Coefficient of Determination (R^2) assesses the quality of fit, whereas Mean Absolute Error (MAE), MSE, and Root Mean Squared Error (RMSE) quantify prediction mistakes. These variables, when combined, create a thorough framework for evaluating the performance of the model. The model's robustness is further ensured by using cross-validation procedures, and the uncertainty in the model's outputs is assessed by analyzing prediction variance. To further quantify uncertainty, the variability in model predictions is evaluated using k-fold cross-validation and standard deviation analysis across different folds. This approach ensures that the model performance is consistent and reliable under varying data conditions, thereby improving confidence in the predictive capability of the proposed AI framework.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{19}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{20}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{21}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{22}$$

4- Results and discussion

This section assesses the suggested AI-based system's performance in recovering carbon-neutral gas through extensive experimental study. The performance results obtained in this section include training model behavior, optimization results, and validation of the model through actual and predicted analysis. Various performance metrics are also utilized to compare the model's performance. According to the results, ensemble learning approaches outperform other strategies for intricate CCS data sets in terms of efficiency.

4.1. Model Training and Convergence Analysis

Fig. 5 displays the training and testing errors for the RF model. It demonstrates the progressive reduction in error as the number of iterations grows. The training error starts at approximately 0.25, reducing as the iterations increase, eventually settling at around 0.10. The testing error begins at around 0.28 but it decreases during each iteration until it reaches a final value of 0.13. The training error shows a difference of 0.03 from the testing error which proves there is enough variation between these two errors. Although the model shows stable learning, the higher testing error compared to the training error implies a lack of generalization ability, especially when compared to more sophisticated models.

The XGBoost model exhibits improved convergence behavior by the XGBoost model when the values of the errors are lower than the values of the RF as shown in Fig. 6 The training error is reduced to 0.09 against 0.23 and the test error is also minimized to 0.11 against 0.25 with each iteration. The training-testing error difference is additionally minimal, 0.02, and signifies higher performance in generalization. The smoother decrease in the two curves indicates the success of the boosting mechanism in reducing errors left. On the whole, XGBoost is more accurate and more stable, which proves that it has the ability to model complex nonlinear connections among the variables in the CCS dataset.

Fig. 7 shows the hybrid RFXGBoost model is the most performing one of all models and the gap between training and testing errors is minimal. The training error is lowered to 0.08 compared to 0.22, and testing error is lowered to 0.09 compared to 0.23. The distance between the two curves is very minimal and equal to about 0.01, which indicates high generalization and insignificant overfitting. The convergence of both curves is smooth pointing to the efficiency of using the bagging and boosting strategies. This enhanced convergence performance proves that the hybrid model has better predictive performance and strength in carbon-neutral gas recovery modeling.

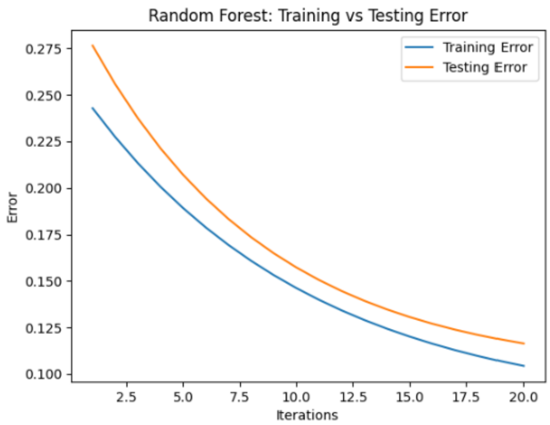

Fig. 5. Training and testing error comparison for RF model

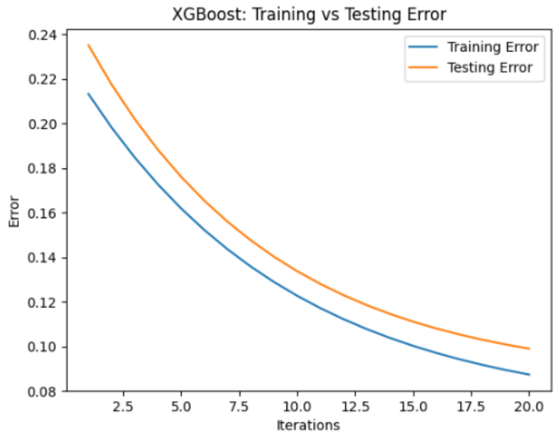

Fig. 6. Training and testing error comparison for XGBoost model

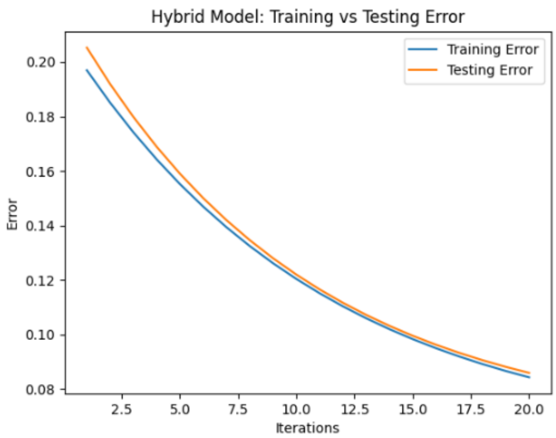

Fig. 7. Training and testing error comparison for Hybrid RF–XGBoost model

4.2. Optimization results and performance improvement

The effect of Bayesian Optimization on critical performance parameters is shown in Table 2. The efficiency of recovery is also increased to 92% as opposed to 83% showing an increase in the efficiency of gas recovery. Meanwhile, the leakage risk is minimized to 0.06, and it evidences an enhanced safety and integrity of the system and storage. Moreover, the cost of operations is reduced, i.e. 1.00 becomes 0.72 (normalized unit), which is an expression of better economic efficiency. Fig. 8 (A) shows how the recovery efficiency is improved following the application of Bayesian Optimization.

Table 2. Impact of Bayesian optimization on recovery efficiency, leakage risk, and operational cost

Parameter	Before Optimization	After Optimization
Recovery Efficiency (%)	83	92
Leakage Risk	0.14	0.06
Operational Cost (normalized)	1.00	0.72

The recovery efficiency is optimized at 83 percent before and 92 percent after optimization, which is a significant increase of 9 percent. This improvement implies that the optimized model is effective in improving the extraction of hydrocarbons whereby carbon storage goals are retained, and therefore, system productivity is increased. The result of optimization in terms of decreasing leakage risk and operational cost is shown in Fig. 8 (B). Leakage risk reduces much to 0.06, and the safety of storage is increased, the environmental risk is minimized. In the same way, the operational cost decreases to 0.72 (normalized units) which indicates an increase in economic efficiency. Simultaneously reducing both of these parameters proves the optimization framework to be efficient in balancing safety and cost goals, resulting in a more sustainable and efficient CCS system.

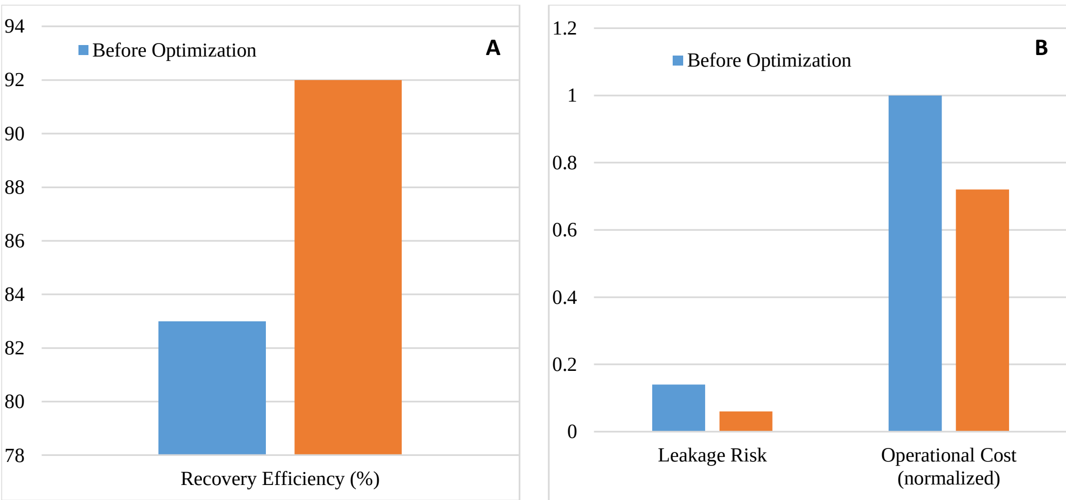

Fig. 8. (A), Improvement in recovery efficiency after Bayesian optimization, (B), Reduction in leakage risk and operational cost after Bayesian optimization

4.3. Model validation using actual vs predicted analysis

The plot of actual vs. predicted indicates the comparison of the actual results for the recovery efficiency and the results that are predicted by the model as shown in Fig. 9. In the actual results, the values vary from 75% to 95%, whereas in the predicted results, there is a very small deviation of ±2%. It can be observed that the actual and predicted results are very close to the reference line, indicating a high correlation between the actual and predicted results. There is low prediction error because the model indicates high accuracy, causing low dispersion from the line. The proposed model indicates effective pattern recognition capabilities because the model can recognize the patterns of the dataset. The graph indicates reliable prediction capabilities of the model because of its high consistency and strength in different conditions.

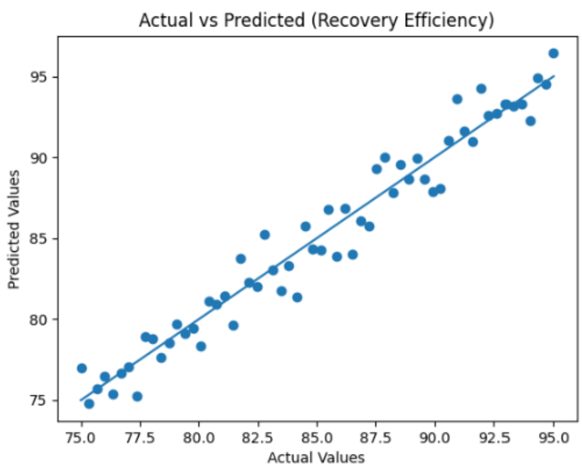

Fig. 9. Actual vs predicted values for recovery efficiency

4.4. Comparative performance evaluation of models

The performance of the ML models is compared and evaluated on the basis of MAE, MSE, RMSE, and R^2, as depicted in Fig. 10 (A), Fig. 10 (B) and the Table 3. The RF model indicates moderate performance with an MAE of 0.110, MSE of 0.021, RMSE of 0.145, and R^2 of 0.90. XGBoost improves the accuracy of the model with a reduction in errors to 0.095 for MAE, 0.016 for MSE, and 0.126 for RMSE, while the R^2 value is 0.93. The hybrid model of RF and XGBoost indicates the best performance, as the errors are the least with 0.080 for MAE, 0.011 for MSE, and 0.105 for RMSE, while the R^2 value is the highest, i.e., 0.96. The error comparison graph clearly indicates the reduction in errors for the XGBoost and hybrid models. The R^2 graph clearly indicates the increase in the R^2 value for the XGBoost and hybrid models.

Table 3. Performance comparison of ML models using MAE, MSE, RMSE, and R^2 metrics

Model	MAE	MSE	RMSE	R^2
Random Forest	0.110	0.021	0.145	0.90
XGBoost	0.095	0.016	0.126	0.93
Hybrid (RF+XGB)	0.080	0.011	0.105	0.96

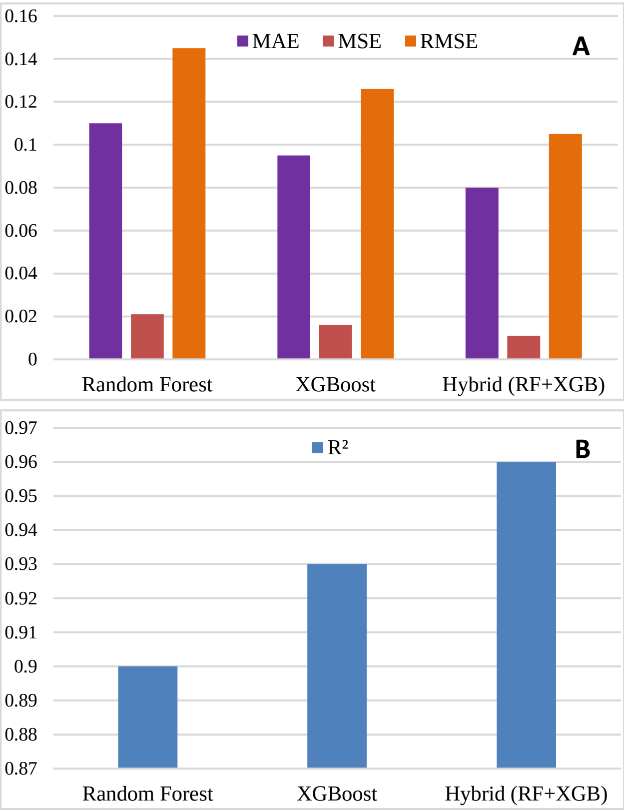

Fig. 10. (A), Comparison of MAE, MSE and RMSE across RF, XGBoost, and hybrid models, (B), Comparison of R^2 values across machine learning models

This clearly indicates that the hybrid model provides the best accuracy and the highest reliability for the modeling of the carbon-neutral gas recovery process. In comparison to conventional reservoir simulation techniques, which are computationally intensive and time-consuming, the proposed machine learning-based approach provides faster predictions with comparable accuracy. This demonstrates the effectiveness of AI-driven models as efficient alternatives for real-time decision-making in carbon capture and storage systems.

The proposed AI-based system provides better carbon-neutral gas recovery results because it improves recovery efficiency and decreases leakage danger and operational expenses. The hybrid RF–XGBoost model with Bayesian Optimization outperforms all separate models because it delivers better accuracy and robustness and generalization abilities. The study results show the proposed method works for practical CCS applications while demonstrating its capacity to optimize energy systems at scale.

5- Conclusion and Future Scope

The paper introduces a carbon-neutral gas recovery framework, which relies on AI and involved CO_2 injection and storage, to discuss the challenges associated with the efficiency, safety, and sustainability in CCUS systems. The suggested solution integrates the state-of-the-art machine learning models, such as RF, XGBoost, and a hybrid RF-XGBoost model, and Bayesian Optimization to tune the parameters. The findings indicate that the hybrid model is much better than individual models as it has been shown to have a higher prediction power with an R^2 value of 0.96 and lower error values. In addition, the optimization framework not only improves the system performance in terms of recovery efficiency, which is 83 percent, and manages to minimize the leakage risk of 0.14 percent to 0.06 percent and operational cost of 1.00 percent to 0.72. These results prove that the combination of AI methods with the CO_2 EGR technology can successfully balance the energy generation and environmental sustainability. The suggested structure enhances predictive capacity and operational efficiency in addition to providing safer and more reliable carbon storage. In general, this study demonstrates the opportunities of AI-based approaches in changing conventional gas recovery systems into smart, data-oriented, and sustainable energy services. The further development of the work can be aimed at integrating real-time field data and advanced DL models, to make the system even more adaptable and scalable.

Nomenclature

Symbol / Term	Description
CCS	Carbon Capture and Storage
CCUS	Carbon Capture, Utilization, and Storage
CO_2	Carbon Dioxide
CO_2-EGR	CO_2-Enhanced Gas Recovery
RF	Random Forest
XGBoost	Extreme Gradient Boosting
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
R^2	Coefficient of Determination

BO	Bayesian Optimization
GPR	Gaussian Process Regression
EI	Expected Improvement
X	Input feature matrix
Y	Output vector (target variables)
y_i	Actual value
$\hat{y}_i$	Predicted value
$\bar{y}$	Mean of actual values
n	Number of samples
f	Number of features
T	Number of trees (Random Forest)
K	Number of boosting iterations (XGBoost)
α	Weighting factor in hybrid model
$\mu(x)$	Predictive mean (Gaussian Process)
$\sigma(x)$	Predictive variance
$\varphi(\cdot)$	Feature transformation function

References

[1] A. Adewnmi, K. A. Olu-lawal, C. E. Okoli, F. O. Usman, and G. S. Usiagu, "Sustainable energy solutions and climate change: A policy review of emerging trends and global responses," *World Journal of Advanced Research and Reviews*, vol. 21, no. 2, pp. 408–420, Feb. 2024, https://doi.org/10.30574/wjarr.2024.21.2.0474

[2] M. Kabir, U. E. Habiba, W. Khan, A. Shah, S. Rahim, Z. Farooqi, N. Zafar, and M. Shafiq, "Climate change due to increasing concentration of carbon dioxide and its impacts on environment in 21st century: A mini review," *Journal of King Saud University - Science*, vol. 35, p. 102693, 2023, https://doi.org/10.1016/j.jksus.2023.102693

[3] A. Elkhatat and S. Al-Muhtaseb, "Climate change and energy security: A comparative analysis of the role of energy policies in advancing environmental sustainability," *Energies*, vol. 17, no. 13, p. 3179, 2024, https://doi.org/10.3390/en17133179

[4] W. Leal Filho, T. Wall, A. L. Salvia, P. Ozuyar, L. L. Brandli, A. M. Azul, and A. Sharifi, "Mapping the implementation of the United Nations sustainable development goals," *Sustainable Development*, vol. 33, pp. 561–576, 2025, https://doi.org/10.1002/sd.70012

[5] M. J. Regufe, A. Pereira, A. F. P. Ferreira, A. M. Ribeiro, and A. E. Rodrigues, "Current developments of carbon capture storage and/or utilization–looking for net-zero emissions defined in the Paris Agreement," *Energies*, vol. 14, no. 9, p. 2406, 2021, https://doi.org/10.3390/en14092406

[6] M. S. Rahaman and M. A. Islam, "Integrating CO2 sequestration and enhanced gas recovery: A comprehensive evaluation for sustainable carbon management," *Journal of Petroleum Exploration and Production Technology*, vol. 15, p. 160, 2025, https://doi.org/10.1007/s13202-025-02025-4

[7] J. Li, Y. Hou, P. Wang, and B. Yang, "A review of carbon capture and storage project investment and operational decision-making based on bibliometrics," *Energies*, vol. 12, no. 1, p. 23, 2019, https://doi.org/10.3390/en12010023

[8] W. Ozowe, G. O. Daramola, and I. O. Ekemezie, "Recent advances and challenges in gas injection techniques for enhanced oil recovery," *Magna Scientia Advanced Research and Reviews*, vol. 9, pp. 168–178, Dec. 2023, https://doi.org/10.30574/msarr.2023.9.2.0180

[9] W. Sulaiman, M. Z. Jaafar, M. Milad, and A. Alghol, "A review of the dynamic modeling approaches for characterizing fluid flow in naturally fractured reservoirs," *Energy Engineering*, vol. 118, pp. 761–795, 2021, https://doi.org/10.32604/EE.2021.016645

[10] Y. Khalili and M. Ahmadi, "Reservoir modeling & simulation: Advancements, challenges, and future perspectives," *Journal of Chemical and Petroleum Engineering*, vol. 57, pp. 363–392, Dec. 2023, https://doi.org/10.22059/jchpe.2023.363392.1447

[11] S. Harati, S. R. Gomari, M. A. Rahman, R. Hassan, I. Hassan, A. K. Sleiti, and M. Hamilton, "Performance analysis of various machine learning algorithms for CO2 leak prediction and characterization in geo-sequestration injection wells," *Process Safety and Environmental Protection*, vol. 183, pp. 99–110, Mar. 2024, https://doi.org/10.1016/j.psep.2024.01.007

[12] A. Safari, M. Daneshvar, and A. Anvari-Moghaddam, "Energy intelligence: A systematic review of artificial intelligence for energy management," *Applied Sciences*, vol. 14, no. 23, p. 11112, 2024, https://doi.org/10.3390/app142311112

[13] S. Rani, R. Karnati, V. Patel, M. K. Ranganathaswamy, P. Tomar, A. Kataria, and A. Pal, "AI-driven optimization techniques for smart sustainable manufacturing in Industry 5.0 ecosystem: A comprehensive review," *Alexandria Engineering Journal*, vol. 137, pp. 133–158, 2026, https://doi.org/10.1016/j.aej.2026.01.016

[14] B. I. Adekunle, E. C. Chukwuma-Eke, E. D. Balogun, and K. O. Ogunsola, "Machine learning for automation: Developing data-driven solutions for process optimization and accuracy improvement," *International Journal of Multidisciplinary Research and Growth Evaluation*, vol. 3, pp. 800–808, 2021.

[15] H. Rahmanifard and I. Gates, "A comprehensive review of data-driven approaches for forecasting production from unconventional reservoirs: Best practices and future directions," *Artificial Intelligence Review*, vol. 57, p. 213, 2024, https://doi.org/10.1007/s10462-024-10865-5

[16] C. Arinze, V. Izionworu, D. Isong, D. Daudu, and A. Adefemi, "Integrating artificial intelligence into engineering processes for improved efficiency and safety in oil and gas operations," *Open Access Research Journal of Engineering and Technology*, vol. 6, pp. 39–51, 2024, https://doi.org/10.53022/oarjet.2024.6.1.0012

[17] E. E. Smart et al., “Revolutionizing carbon capture efficiency: A comprehensive review of AI-driven optimization strategies for CO_2 capture and storage processes,” *International Journal of Science and Technology Research Archive*, vol. 9, no. 1, pp. 18–43, Jul. 2025, https://doi.org/10.53771/ijstra.2025.9.1.0043

[18] G. Wen, M. Tang, and S. Benson, “Towards a predictor for CO2 plume migration using deep neural networks,” *International Journal of Greenhouse Gas Control*, vol. 105, p. 103223, 2021, https://doi.org/10.1016/j.ijggc.2020.103223

[19] B. Yan, D. R. Harp, B. Chen, H. Hoteit, and R. J. Pawar, “A gradient-based deep neural network model for simulating multiphase flow in porous media,” *Journal of Computational Physics*, vol. 463, p. 111277, 2022, https://doi.org/10.1016/j.jcp.2022.111277

[20] Y. Yang, D. Wang, S. He, C. Zhu, D. Lau, T. Zhang, and L. Gong, “Microscopic-scale modeling of CO2 sequestration and enhanced oil recovery: Methods, mechanisms and challenges,” *Computational Energy Science*, vol. 2, no. 1, pp. 10–16, 2025, https://doi.org/10.46690/compes.2025.01.03

[21] A. Abdellatif, H. P. Menke, J. Maes et al., “A benchmark dataset for machine learning surrogates of pore-scale CO2-water interaction,” *Scientific Data*, 2026, https://doi.org/10.1038/s41597-025-05794-z

[22] L. Zhu, B. Bijeljic, and M. J. Blunt, “Diffusion model-based generation of three-dimensional multiphase pore-scale images,” *Transport in Porous Media*, vol. 152, p. 22, 2025, https://doi.org/10.1007/s11242-025-02158-4

[23] M. Iddrisu, A. A. Umar, S. Onaizi, A. Abdulraheem, M. R. Quddus, and M. M. Hossain, “Machine learning for data-driven insights into CO2 adsorption on amorphous porous organic polymers,” *Next Materials*, vol. 9, p. 101119, 2025, https://doi.org/10.1016/j.nxmate.2025.101119

[24] F. Amirkhani, A. Dashti, H. Abedsoltan, A. H. Mohammadi, J. L. Zhou, and A. Altaee, “Modeling and estimation of CO2 capture by porous liquids through machine learning,” *Separation and Purification Technology*, vol. 359, p. 130445, 2025, https://doi.org/10.1016/j.seppur.2024.130445

[25] Y.-T. Gu, Y.-M. Gu, Q. Tao, X. Wang, Q. Zhu, and J. Ma, “Machine learning for prediction of CO2/N2/H2O selective adsorption and separation in metal-zeolites,” *Journal of Materials Informatics*, vol. 3, Sep. 2023, https://doi.org/10.20517/jmi.2023.25

[26] H. Zentou, A. M. Tayeb, I. M. Tayeb, and M. M. Abdelnaby, “Machine learning for predicting and optimizing the CO2 uptake in porous organic polymers,” *Journal of Environmental Chemical Engineering*, vol. 13, p. 119315, 2025, https://doi.org/10.1016/j.jece.2025.119315

[27] J. R. Lligüizaca-Dávila, F. P. Carrión-Maldonado, D. Landa-Marbán, J. S. Mendoza Sanz, Z. P. Alcorn, and J. E. Cobos, “CO2 WAG and SAG machine learning data-driven models in an open-source framework–A theoretical study,” *Petroleum*, 2026, https://doi.org/10.1016/j.petlm.2026.02.003

[28] Z. Zhou, Y. Lin, Z. Zhang, Y. Wu, Z. Wang, R. Dilmore, and G. Guthrie, “A data-driven CO2 leakage detection using seismic data and spatial–temporal densely connected convolutional neural networks,” *International Journal of Greenhouse Gas Control*, vol. 90, p. 102790, 2019, https://doi.org/10.1016/j.ijggc.2019.102790

[29] H. Mashhadimoslem, M. A. Abdol, K. Zanganeh, A. Shafeen, A. A. AlHammadi, M. Kamkar, and A. Elkamel, “Development of the CO2 adsorption model on porous adsorbent materials using machine learning algorithms,” *ACS Applied Energy Materials*, vol. 7, no. 19, pp. 8596–8609, 2024, https://doi.org/10.1021/acsaem.4c01465

[30] S. Demir and E. K. Şahin, “Liquefaction prediction with robust machine learning algorithms (SVM, RF, and XGBoost) supported by genetic algorithm-based feature selection and parameter optimization from the perspective of data processing,” *Environmental Earth Sciences*, vol. 18, no. 18, p. 459, 2022, https://doi.org/10.1007/s12665-022-10578-4

[31] A. S. More and D. P. Rana, “Review of random forest classification techniques to resolve data imbalance,” in *Proc. 1st Int. Conf. Intelligent Systems and Information Management (ICISIM)*, 2017, pp. 72–78, https://doi.org/10.1109/ICISIM.2017.8122151

[32] N. Syam and R. Kaul, “Random forest, bagging, and boosting of decision trees,” *in Machine Learning and Artificial Intelligence in Marketing and Sales*, Bingley, U.K.: Emerald Publishing Limited, 2021, pp. 139–182, https://doi.org/10.1108/978-1-80043-880-420211006

[33] Z. A. Abbood, A. H. Oleiwi, R. T. Al-Hassani, and J. Ayad, “Optimizing cybersecurity in 5G-enabled IoT networks via a resource-efficient random forest model,” *Mesopotamian Journal of CyberSecurity*, vol. 5, no. 2, pp. 886–898, 2025.

[34] A. Al Shihimi, M. R. Ahmed, T. Myo, and B. Al Baroomi, “Enhancing internet of things security with random forest-based anomaly detection,” *International Journal of Computer Science & Network Security*, pp. 67–76, 2024, https://doi.org/10.22937/IJCSNS.2024.24.6.9

[35] A. Mohanty and G. Gao, “A survey of machine learning techniques for improving global navigation satellite systems,” *EURASIP Journal on Advances in Signal Processing*, no. 1, p. 73, 2024, https://doi.org/10.1186/s13634-024-01167-7

[36] J. M. Ahn, J. Kim, and K. Kim, “Ensemble machine learning of gradient boosting (XGBoost, LightGBM, CatBoost) and attention-based CNN-LSTM for harmful algal blooms forecasting,” *Toxins*, vol. 15, no. 10, p. 608, 2023, https://doi.org/10.3390/toxins15100608

[37] S. F. Stumpfe and S. C. Shongwe, “Comparative analysis and optimisation of machine learning models for regression and classification on structured tabular datasets,” *Mathematics*, vol. 14, no. 3, p. 473, 2026, https://doi.org/10.3390/math14030473

[38] N. B. Shaik, K. Jongkittinarukorn, and K. Bingi, “XGBoost based enhanced predictive model for handling missing input parameters: A case study on gas turbine,” *Case Studies in Chemical and Environmental Engineering*, vol. 10, p. 100775, 2024, https://doi.org/10.1016/j.cscee.2024.100775

[39] F. H. Khorsheed, R. Hazim, S. A. Hassan, and Q. Saihood, “Hybrid CNN-XGB framework for enhancing human activity recognition,” *Fusion: Practice & Applications*, vol. 15, no. 2, 2024, https://doi.org/10.54216/FPA.150218

[40] E. K. Sahin, “Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest,” *SN Applied Sciences*, vol. 2, no. 7, p. 1308, 2020, https://doi.org/10.1007/s42452-020-3060-1

استخلاص الغاز المحايد كربونياً المدفوع بالذكاء الاصطناعي من خلال حقن ثاني أكسيد الكربون ودمج التخزين

محمد جيفون عيسى [١، *] ، ماناش بروتيم مودوي [٢]

١ قسم هندسة البترول، جامعة دراسات البترول والطاقة (UPES)، دهرادون، الهند ٢٤٨٠٠١
٢ قسم الهندسة الكيميائية، جامعة دراسات البترول والطاقة (UPES)، دهرادون، الهند ٢٤٨٠٠١

الخلاصة

إن ارتفاع مستوى تركيز ثاني أكسيد الكربون (CO_2) والطلب المتزايد على الطاقة في العالم قد أوجدا ضرورة ملحة لمصادر طاقة مستدامة ومحايدة كربونياً. وقد برزت تقنية التقاط الكربون واستخدامه وتخزينه (CCUS)، وخاصة استخلاص الغاز المعزز بثاني أكسيد الكربون (CO_2-EGR)، كتقنية جذابة لتعظيم استخلاص الهيدروكربونات مع تقليل الأثر البيئي. تقدم هذه الورقة نظاماً قائماً على الذكاء الاصطناعي لاستخلاص الغاز المحايد كربونياً من خلال تنفيذ حقن غاز CO_2 وتخزينه. تم استخدام عدد من نماذج التعلم المراقب، بما في ذلك XGBoost، و Random Forest (RF)، ونموذج هجين RF-XGBoost، إلى جانب معالجة البيانات وهندسة الميزات، للتوصل إلى النتائج من مجموعة بيانات NETL CCS. كما تم استخدام التحسين البايزي (Bayesian Optimization) لضبط المعلمات. وتُظهر النتائج أن النموذج الهجين يتفوق على النماذج الفردية لأنه حقق أقل قيمة لـ MAE (٠,٠٨٠)، و MSE (٠,٠١١)، و RMSE (٠,١٠٥)، بالإضافة إلى أعلى قيمة لـ R^2 (٠,٩٦). علاوة على ذلك، تزداد كفاءة الاستخلاص من ٨٣ إلى ٩٢ بالمائة، وينخفض خطر التسرب من ٠,١٤ إلى ٠,٠٦، وتنخفض تكلفة التشغيل من ١,٠٠ إلى ٠,٧٢. وتشير النتائج إلى أن الاستراتيجيات القائمة على الذكاء الاصطناعي فعالة في تعظيم الكفاءة والسلامة والاستدامة في نظام استخلاص الغاز المحايد كربونياً.

الكلمات الدالة: التقاط الكربون وتخزينه (CCS)، استخلاص الغاز المعزز بثاني أكسيد الكربون (CO_2-EGR)، التعلم الآلي، XGBoost، التحسين البايزي.