Engineering Technology Trends Vol 3 Issue 1 2025

Original Research Articles

Open Access

Exploring the Application of Machine Learning in Carbon Capture, Utilization, and Storage Technologies

Bo Zhang[1,2,*] **Chuan Zhang**[1]

[1]School of Economics and Management, Shanghai University of Electric Power, Shanghai, China, 201306.
[2]CAS Key Laboratory of Low-Carbon Conversion Science & Engineering, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, China, 201210.

***Correspondence to:** Bo Zhang, School of Economics and Management, Shanghai University of Electric Power, Shanghai, China, 201306. Email: zhangbo2024@sari.ac.cn.

Abstract: Carbon capture, utilization, and storage (CCUS) technologies are key solutions to mitigating climate change. In recent years, with the advancement of technology, the application of machine learning (ML) in optimizing various stages of CCUS technologies has garnered increasing attention. This paper summarizes the current status of ML applications in CO_2 capture, CO_2 enhanced oil recovery, CO_2 storage, underground sequestration, as well as in the evaluation of the CCUS technologies chain and source-sink matching. It explores the use of ML theory to optimize processes such as data collection, model prediction and recognition, model parameter adjustment, and result comparison in relation to CCUS technologies. Additionally, a full-chain model for CCUS technologies is constructed, and the future directions of ML in CCUS are envisioned. The research provides insights to fully harness the potential of machine learning in the CCUS field.
Keywords: Carbon Capture Utilization and Storage (CCUS) technologies; Machine learning

1. Introduction

Carbon capture, utilization, and storage (CCUS) technologies have emerged as key strategies for addressing climate change and have garnered widespread attention in recent years. With the continuous development of CCUS technologies, machine learning (ML) has increasingly become an essential tool for optimizing the application of these technologies. ML focuses on improving system performance by utilizing computational methods and experiential data. Significant potential has been demonstrated in optimizing CO_2 capture efficiency, CO_2 enhanced oil recovery (EOR), as well as in monitoring underground sequestration and assessing safety. The construction of CCUS technologies models involves various types of data, including, but not limited to, carbon emission intensity and total emissions across different geographical regions and time scales, the economic evaluation of various CCUS technologies, and the estimation of CO_2 disposal capacity. These data often exhibit complex nonlinear relationships and spatiotemporal variability, which require advanced ML algorithms for intelligent processing and modeling. However, issues such as data quality, integration of

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

physical and ML models, computational resources, and model transparency remain key challenges limiting further development.

This paper explores the current status and future trends of ML applications in CCUS technologies, considering how ML theories can optimize processes related to data collection, model prediction and recognition, parameter adjustment, and result comparison. The goal is to further enhance efficiency, accuracy, and intelligence, thereby promoting innovation and optimization in CCUS technologies. Additionally, the paper looks forward to the role of ML in driving technological advancements in CCUS and addressing climate change, providing both theoretical foundations and practical guidance for the future intelligent optimization of CCUS technologies.

2. The current application of ML in CCUS technologies

The unique advantages of machine learning (ML), especially in handling complex nonlinear relationships, optimizing multi-dimensional variables, and uncovering hidden patterns in large-scale datasets, have shown great promise across various stages of CCUS. Based on a review of recent literature, the application of ML in CCUS can be summarized in several key areas.

Firstly, in CO_2 capture processes, ML is primarily focused on optimizing capture efficiency and predicting key parameters in the capture process. For instance, supervised learning techniques are used to model the physicochemical properties of adsorbents and solvents, effectively predicting performance indicators such as adsorption capacity, separation efficiency, and energy consumption during the capture process[1]. Recently, deep learning and neural network-based techniques have been applied to predict CO_2 behavior in various adsorbent materials, further enhancing the accuracy and efficiency of the capture process. Additionally, surrogate models have emerged as a significant research direction in CO_2 capture[2]. By training these models using ML, high prediction accuracy can be maintained while reducing computational costs, thereby facilitating efficient optimization of the CO_2 capture process.

Secondly, in CO_2-enhanced oil recovery (CO_2-EOR), ML methods have also found widespread applications. By analyzing large-scale subsurface exploration data, production data, and historical operational data, ML models can effectively predict the characteristics of oil and gas reservoirs and CO_2 injection behavior[3]. Algorithms such as Support Vector Machines (SVM) and Random Forests (RF) have been used in tasks like reservoir modeling, CO_2 injection optimization, and production forecasting. Furthermore, ML has been applied to real-time monitoring of the CO_2 injection process, enabling the identification of potential issues based on downhole pressure, temperature, and other data, thus optimizing injection strategies and improving the economic efficiency of CO_2-EOR[4].

Thirdly, in the fields of CO_2 storage and underground sequestration, ML is primarily focused on CO_2 migration, storage safety assessment, and leakage detection. The flow and distribution of CO_2 in underground reservoirs are influenced by various factors, and traditional numerical simulation methods often struggle to rapidly and accurately reflect the complexities of subsurface environments. ML, especially models like Deep Neural Networks (DNN) and Long Short-Term Memory (LSTM) networks, has demonstrated strong capabilities in analyzing subsurface fluid flow and predicting CO_2 migration paths[5]. By performing real-time analysis of seismic data and downhole monitoring data (e.g., pressure, temperature, porosity), ML can identify potential leakage pathways and risks, ensuring the safety of the sequestration process. In recent years, ML methods combined with Bayesian inference have also made significant progress in uncertainty analysis during sequestration, further improving model reliability and predictive accuracy[6].

Finally, ML has shown remarkable potential in the evaluation and source-sink matching of CCUS technologies chains. By integrating multi-source data and applying pattern recognition, ML can optimize the matching process between sources and storage sites, improving matching efficiency and accuracy, particularly when dealing with complex nonlinear relationships and dynamic environments[7]. The application of deep learning and reinforcement learning has further advanced precise prediction and adaptive optimization, supporting multi-objective optimization, risk assessment, and dynamic performance adjustments. In the evaluation of the entire CCUS technologies chain, ML can help achieve goals such as cost minimization and maximization of sequestration efficiency, while also providing quantitative risk assessments and continuous monitoring of system performance, significantly enhancing the deployment efficiency of CCUS across various steps in the value chain[8]. As data acquisition technologies and algorithms advance, ML is expected to play an increasingly critical role in the CCUS field, driving the global achievement of emission reduction targets. However, challenges such as data quality, algorithm transparency, and interpretability need to be addressed to ensure the reliability and acceptability of widespread applications.

Despite the numerous advantages of ML in CCUS, its application still faces several challenges. Firstly, issues related to data quality and missing data are particularly prominent in CCUS, especially since underground data acquisition is constrained by monitoring technology limitations, which can lead to data bias and inaccuracies during model training. Secondly, effectively integrating physical models with ML algorithms to create interdisciplinary hybrid models remains a critical challenge. Physical models provide the theoretical foundation for system behavior, while ML discovers hidden patterns through data-driven approaches. The fusion of these two approaches holds the potential to significantly enhance model prediction capabilities and decision-support outcomes. Finally, the computational complexity and interpretability issues of ML models are key challenges in current applications. In large-scale, multi-dimensional CCUS tasks, the complex model training and inference processes require substantial computational resources, while the "black box" nature of these models makes them difficult to interpret and validate in practical applications.

3. ML-Based full Chain model of CCUS technologies

In ML applications, there is typically a discrepancy between the model's predicted output and the true sample values. Minimizing this discrepancy is crucial for ensuring that the model accurately reflects the actual process. In the application of ML theory to carbon emission data, model prediction optimization, carbon dioxide sequestration system construction, and CCUS source-sink matching, errors between the model's output and the actual data are inevitable.

First, the error of the ML model must be experimentally verified, followed by an evaluation and selection process. Evaluation metrics include error rate, accuracy, precision, recall, and others. Second, for CCUS technologies applications in key emission-intensive industries, training models based on historical "processing experience" data can further improve prediction accuracy. Given the complexity of CCUS source-sink matching, a multi-class learning approach can be adopted, breaking the problem into several binary classification tasks, where each task trains an independent classifier, and the final classification decision is obtained by aggregating the predictions of the individual classifiers.

Furthermore, neural network models can be trained on pre-processed data, adjusting the network's weights and biases through multiple iterations to minimize the loss function. In this process, evaluating the model's performance using test data is an indispensable step. Based on the evaluation results, the model can be optimized, including adjustments to hyperparameters such as the number of hidden layers, neurons, activation functions, and optimization algorithms. Additionally, to prevent model overfitting, regularization techniques and dropout methods can be applied effectively. For the dynamic factors in the source-sink matching process (such as fluctuations in source production and adjustments in sink capacity), the model's performance should be periodically assessed, and necessary adjustments should be made based on these changes.

In the process of using CCUS technologies to reduce carbon emissions in key industries, the first step is to collect relevant data on carbon emission intensity and total emissions based on the production process, as well as internal and remediation cost data related to carbon emissions, and the absorption capacity and related costs of various carbon sinks. These data types are numerous and often incomplete or missing, and manual calculations are time-consuming and resource-intensive. Therefore, based on a ML framework, the raw data should first undergo preprocessing, including noise elimination, missing value imputation, and data normalization. Subsequently, supervised learning methods should be used to deeply analyze the existing data and construct and optimize carbon emission source-sink matching models and carbon dioxide sequestration systems for key industries. During the evaluation phase, the system's performance will be comprehensively assessed using multidimensional metrics such as accuracy, precision, recall, F1 score, and AUC-ROC curve. To further improve the model's performance, advanced techniques like grid search, random search, and Bayesian optimization will be used to optimize the model's hyperparameters. Finally, the trained CCUS source-sink matching model will be deployed in a real production environment to support the effective implementation and optimization of CCUS projects.

Through the above process, the application effectiveness of CCUS technologies in emissions reduction can be enhanced to some extent, promoting the optimization of carbon emission management and carbon dioxide sequestration in key industries. **Figure 1**. illustrates the conceptual diagram of the full-chain model for carbon capture, utilization, and storage (CCUS) based on ML.

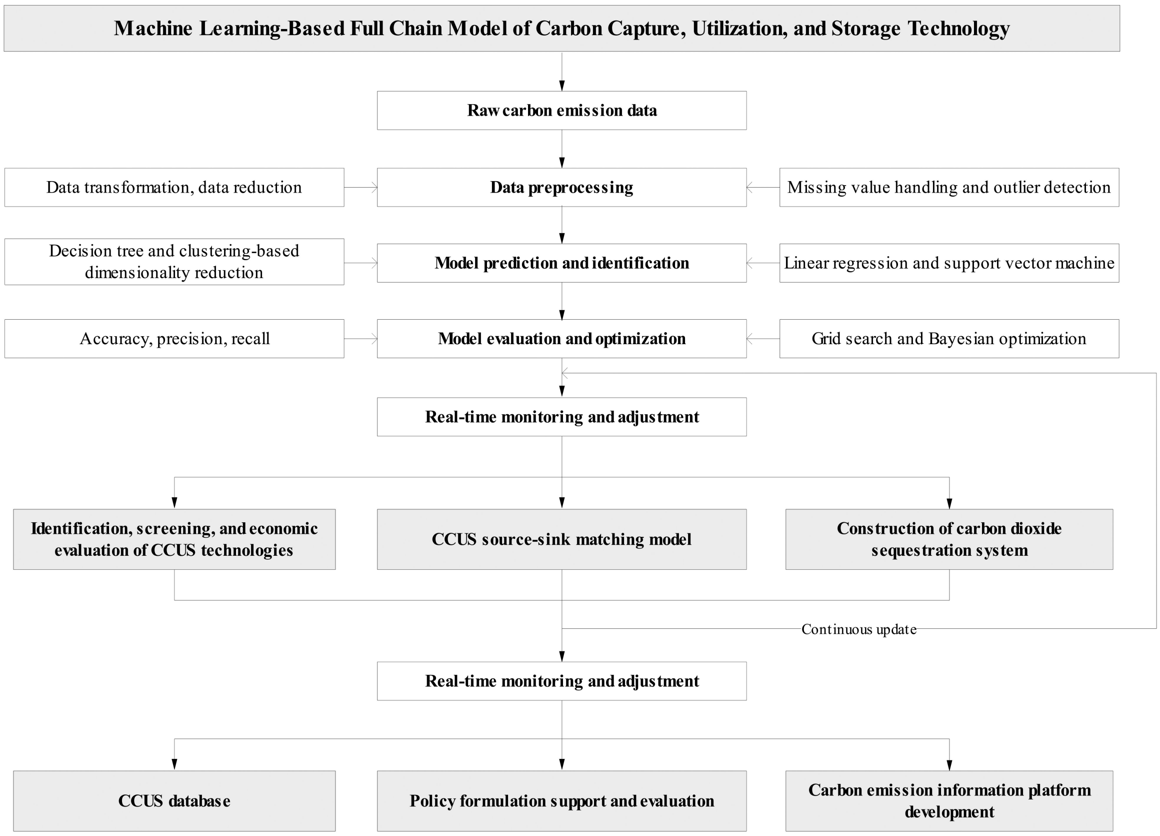

Figure 1. ML-Based Full Chain Model of CCUS Technologies

4. Development directions of ML in CCUS technologies

The application of ML technologies in CCUS is advancing rapidly and demonstrating significant potential across various domains. As global efforts to combat climate change intensify, CCUS has emerged as a key emissions reduction technology, and optimizing its various stages requires the aid of advanced data-driven methods. In the future, the research directions for ML in CCUS technologies are expected to focus on several key areas:

First, interdisciplinary data fusion and multi-source information integration will become central. ML has the ability to integrate data from various fields, such as geology and environmental science, thereby improving the accuracy of source-sink matching and risk assessment. Second, real-time dynamic optimization and adaptive control will play a pivotal role in source-sink matching and transportation path optimization. Reinforcement learning and online learning methods are expected to achieve autonomous optimization of the system through interaction with real-time data streams.

Third, deep learning will be widely applied in system-level integration of CCUS. Models such as deep neural networks can integrate multi-stage data to provide optimized solutions across the entire process and handle complex nonlinear constraints. At the same time, the development of Explainable AI (XAI) will enhance the transparency and comprehensibility of ML models, ensuring their reliability and societal acceptability in decision-making processes. Collaborative optimization and multi-objective decision-making will further improve the collaborative efficiency of various stages within the CCUS system. ML will take into account multiple objectives such as cost and sequestration effectiveness, offering balanced solutions for different stakeholders.

Moreover, long-term evolution and system reliability prediction will become significant areas of research. Time-series deep learning methods will be employed to predict long-term changes and potential risks in storage sites. Finally, carbon market and policy-driven model optimization will provide strategic decision support for CCUS systems within the framework of carbon pricing and carbon trading mechanisms.

In summary, ML will play a crucial role in the future of CCUS technologies. Through intelligent optimization and prediction, ML will drive the efficient implementation and sustainable development of CCUS technologies, providing strong support for achieving global emissions reduction targets.

5. Conclusion

The application of ML in CCUS technologies is gradually deepening, offering robust support for optimizing various stages of the CCUS process. In areas such as CO_2 capture, CO_2-EOR (Enhanced Oil Recovery), and underground storage, ML has enhanced system efficiency and predictive accuracy through data-driven approaches. With the integration of multi-source data and interdisciplinary fusion, ML will further improve the accuracy of source-sink matching and the reliability of safety assessments. Additionally, the application of deep learning, reinforcement learning, and other techniques will drive real-time dynamic optimization and system-level integration of CCUS technologies, enabling better responses to complex climate change challenges.

Despite these advancements, issues such as data quality, model integration, and computational resources remain challenges. Future research should focus on addressing these technical bottlenecks while enhancing the interpretability and transparency of ML models to ensure their broad acceptance and reliability in practical applications. Overall, ML will play a vital role in the development of CCUS technologies, providing critical support for achieving global emissions reduction goals and promoting the transition to a low-carbon economy.

Reference

[1] BHAGAT S K, PILARIO K E, BABALOLA O E, et al. Comprehensive review on machine learning methodologies for modeling dye removal processes in wastewater[J]. Journal of Cleaner Production, 2023, 385: 135522.

[2] TREVISAN L, CIHAN A, FAGERLUND F, et al. Investigation of mechanisms of supercritical CO_2 trapping in deep saline reservoirs using surrogate fluids at ambient laboratory conditions[J]. International Journal of Greenhouse Gas Control, 2014, 29: 35-49.

[3] YAN B, CHEN B, HARP D R, et al. A robust deep learning workflow to predict multiphase flow behavior during geological CO_2 sequestration injection and Post-Injection periods[J]. Journal of Hydrology, 2022, 607: 127542.

[4] ZHONG Z, SUN A Y, YANG Q, et al. A deep learning approach to anomaly detection in geological carbon sequestration sites using pressure measurements[J]. Journal of Hydrology, 2019, 573: 885-894.

[5] WEN G, TANG M, BENSON S M. Towards a predictor for CO_2 plume migration using deep neural networks[J]. International Journal of Greenhouse Gas Control, 2021, 105: 103223.

[6] CHEN B, HARP D R, LIN Y, et al. Geologic CO_2 sequestration monitoring design: A machine learning and uncertainty quantification based approach[J]. Applied Energy, 2018, 225: 332-345.

[7] ALEXANDER F J, ANG J, BILBREY J A, et al. Co-design center for exascale machine learning technologies (exalearn)[J]. The International Journal of High Performance Computing Applications, 2021, 35(6): 598-616.

[8] YAN Y, BORHANI T N, SUBRAVETI S G, et al. Harnessing the power of machine learning for carbon capture, utilisation, and storage (CCUS)–a state-of-the-art review[J]. Energy & Environmental Science, 2021, 14(12): 6122-6157.