European Journal of Sustainable Development Research
2026, 10(2), em0378
e-ISSN: 2542-4742
https://www.ejosdr.com/

MODESTUM

Research Article

OPEN ACCESS

Hybrid machine learning and computational fluid dynamics framework for optimizing supercritical CO_2 pipeline networks in large-scale carbon capture and storage

Abiola Samuel Ajayi [1*], Ayoola David Bodude [2], Shokenu Emmanuel Segun [3], Eniola Isaac Ajayi [4], Dawudu Ibrahim Shanzhi [5]

[1] Department of Computer Science, Texas Southern University, Houston, TX, USA
[2] Department of Chemical Engineering, University of Strathclyde, Galsgow, UK
[3] Department of System Engineering, African University of Science and Technology, Abuja, NIGERIA
[4] Department of Computer Science, Aptech Vocational Training Institute, Abuja, NIGERIA
[5] Department of System Engineering, African University of Science and Technology, NIGERIA
***Corresponding Author:** ajayisamuel645@gmail.com

Citation: Ajayi, A. S., Bodude, A. D., Segun, S. E., Ajayi, E. I., & Shanzhi, D. I. (2026). Hybrid machine learning and computational fluid dynamics framework for optimizing supercritical CO_2 pipeline networks in large-scale carbon capture and storage. *European Journal of Sustainable Development Research, 10*(2), em0378. https://doi.org/10.29333/ejosdr/17977

ARTICLE INFO

Received: 03 Jul. 2025

Accepted: 06 Dec. 2025

ABSTRACT

This study presents an integrated framework combining computational simulations and machine learning (ML) to design and optimize supercritical carbon dioxide (CO_2) pipeline networks for large-scale carbon capture and storage (CCS) applications. Using ANSYS fluent, the fluid dynamics of supercritical CO_2 transport were analyzed under various operational conditions to generate synthetic insights and complement real-world datasets collected from CO_2 sequestration activities and industrial sources. Key variables such as pressure, temperature, and pipeline diameter were extracted, engineered into time-lagged and rolling features, and used to train a random forest regressor model. Feature importance analysis revealed that short-term flow trends, volatility, and geographic segmentation are dominant predictors of CO_2 flow behavior. The ML model achieved high performance, with an R^2 score of 0.9897 and a low RMSE of 0.0335, indicating strong predictive reliability. Hyperparameter tuning further optimized model accuracy, while visual analyses including actual vs. predicted comparisons, residual distributions, and feature interaction heatmaps validated the model's robustness. The results demonstrate that data-driven approaches can complement simulation-based methods to improve the design, monitoring, and operational efficiency of CO_2 transport systems. This work provides a scalable, intelligent solution for CCS pipeline optimization, supporting the global transition toward low-carbon infrastructure.

Keywords: supercritical CO_2, carbon capture and storage, pipeline optimization, ANSYS fluent, machine learning, flow rate prediction, feature engineering, hyperparameter tuning, CO_2 emissions forecasting

INTRODUCTION

The urgency to combat global climate change has accelerated the development and deployment of carbon capture and storage (CCS) technologies. Among the key infrastructure requirements for CCS is the safe, efficient, and cost-effective transportation of carbon dioxide (CO_2) often in its supercritical state from capture facilities to geological storage sites. Designing optimized pipeline networks for such transport requires accurate modeling of CO_2 flow dynamics under varying thermophysical and environmental conditions. To this end, recent advancements in computational fluid dynamics (CFD) and machine learning (ML) have shown considerable promise in improving the efficiency of engineering design workflows for energy-related systems (Alatyar, 2023; Romano, 2023).

Supercritical CO_2 behaves in a non-ideal manner, particularly under high-pressure and temperature regimes, making its transport modeling highly complex. Traditional CFD methods, while accurate, are often computationally expensive and time-intensive when applied to large-scale network simulations. To address this challenge, researchers have begun integrating neural network-based surrogate models to reduce simulation times without sacrificing accuracy (Sahranavardfard, 2024). These AI-driven approaches enable fast evaluation of two-phase, real-fluid thermophysical properties, enhancing the adaptability of simulations for pipeline network design. Surrogate models, when trained on high-fidelity CFD data, can be particularly effective in

predicting flow parameters such as pressure drop, temperature change, and velocity distribution.

In recent studies, open-source CFD solvers and hybrid simulation tools have been applied to optimize complex fluid transport systems, including supercritical CO_2 applications (Qi, 2023; Wang et al., 2024). Qi (2023), for instance, demonstrated the feasibility of combining CFD with AI for the simulation of radial inflow turbines using supercritical CO_2, a concept translatable to pipeline design in CCS infrastructure. Similarly, Wang et al. (2024) highlighted the use of digital twins and control systems in the modeling and simulation of small modular reactors indicating a broader shift toward intelligent design paradigms in thermofluid systems.

Moreover, efforts in the Python-based modeling of CO_2 pipelines and wells have further simplified the integration of ML models with simulation data (Arntzen, 2024). Such integration enables real-time design optimization, helping engineers balance trade-offs between cost, safety, and environmental performance. The significance of scalable, intelligent design tools becomes even more pronounced when dealing with multi-node pipeline networks, where the behavior of CO_2 can vary significantly due to terrain, pipe material, and insulation factors. Udemu (2024) also emphasized the importance of scale-up modeling in process intensification, reinforcing the necessity of cross-disciplinary approaches that combine chemical engineering principles with AI-enhanced simulations.

In addition to the physical modeling, experimental and numerical analyses of fluid networks such as those involving carbonized biomass gasification have helped validate ML-based predictive models (Kuttin et al., 2024). These studies provide essential benchmark data to support ML training and calibration for pipeline optimization. ML algorithms such as random forests, artificial neural networks (ANN), and genetic algorithms (GA) have demonstrated significant capability in identifying patterns, predicting system behavior, and optimizing designs across large solution spaces especially in safety-critical domains like CCS.

This study builds upon these foundations by presenting a hybrid CFD-ML framework for the design and optimization of supercritical CO_2 pipeline networks. Using ANSYS fluent for high-resolution CFD simulation and Python-based ML libraries for predictive modeling, the work aims to enhance the performance, safety, and economic viability of pipeline networks in CCS deployments. Previous efforts by Segun et al. (2023), although in a different application domain, show the potential of RFID and automation in engineering system design and monitoring. This reinforces the broader applicability of AI in solving infrastructural challenges. The paper thus contributes to ongoing discourse on intelligent infrastructure design, offering a scalable, data-driven solution for one of the most critical challenges in CCS implementation.

LITERATURE REVIEW

The urgent need to decarbonize global energy systems has positioned CCS as a pivotal technology in addressing climate change. CCS involves capturing CO_2 from major emission sources, compressing it into a supercritical state, transporting

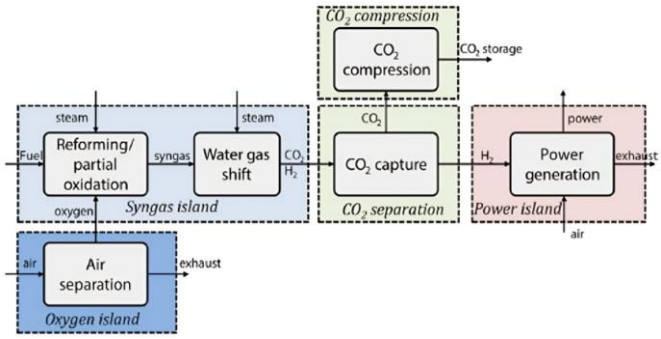

Figure 1. A schematic of a pre-combustion CO_2 capture system illustrating air separation, syngas processing, CO_2 capture, compression, and power generation (this setup captures CO_2 at high pressure and routes it efficiently for compression and storage) (Adapted from Jansen et al., 2015)

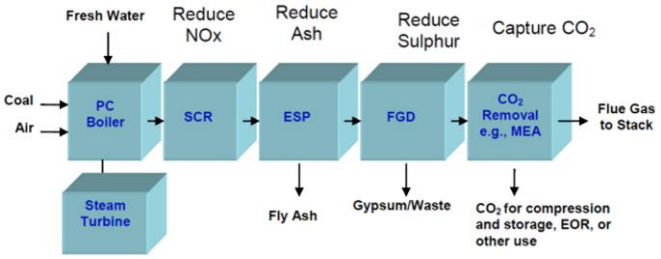

Figure 2. Diagram showing the post-combustion capture of CO_2 in a conventional pulverized coal (PC) power plant (it includes NO_x, ash, and sulfur removal stages before flue gas enters the CO_2 capture unit) (Global CCS Institute, 2012)

it through pipeline networks, and injecting it into underground reservoirs for long-term storage. The two diagrams above reflect typical CCS integration points: the first (from gasification processes) illustrates syngas processing and CO_2 compression, while the second (from coal-fired plants) shows sequential treatment and capture of CO_2 post-combustion. These systems form the operational foundation for CO_2 pipeline transport optimization and have gained renewed attention due to innovations in CFD modeling and AI-driven design (Global CCS Institute, 2012; Jansen et al., 2015) (**Figure 1**).

To accurately model and optimize these complex CCS networks, CFD tools like ANSYS fluent are widely used to simulate supercritical CO_2 flow, predict thermodynamic behavior, and evaluate pressure loss, temperature gradients, and potential leakage risks. However, large-scale simulations often demand significant computational resources. Researchers have begun incorporating ML to develop surrogate models that approximate CFD results more efficiently. These artificial intelligence (AI)-enhanced models accelerate pipeline design cycles and support sensitivity analysis across various topographies and operating conditions (Ajayi et al., 2025; Christianah et al., 2025). ML tools like ANN and GA have been particularly effective in handling nonlinear fluid dynamics within CCS pipeline systems (**Figure 2**).

Moreover, post-combustion and pre-combustion capture technologies differ significantly in how they influence pipeline input conditions. As seen in the second diagram, post-combustion methods (e.g., monoethanolamine scrubbing)

typically deal with low-pressure, flue gas-derived CO_2, requiring substantial energy input for compression. Conversely, pre-combustion systems, as shown in the first diagram, deliver CO_2 at higher purity and pressure, which is better suited for transport (Jansen et al., 2015). Understanding these upstream variations is essential for setting pipeline boundary conditions during ANSYS fluent simulations and for training ML models to generalize effectively across pipeline scenarios.

Recent studies have also explored hybrid energy systems where CCS is integrated with renewable energy or desalination systems to enhance overall efficiency. For example, Traisak et al. (2025) review advancements in thermoelectric power generation coupled with water desalination, which parallels the integration complexity seen in CCS networks. These interdisciplinary systems benefit from AI-driven optimization algorithms that can simultaneously consider energy flow, material costs, and CO_2 transport behavior. Such optimization mirrors the objectives of CCS pipeline design, where pressure loss, risk zones, route distance, and insulation strategy must all be concurrently optimized.

Furthermore, advanced knowledge systems and decision support platforms are beginning to complement traditional design methods. Ukah et al. (2024) presented a knowledge-based support system for infrastructure projects, suggesting that combining expert systems with ML could further streamline CCS design decision-making. Similarly, Ajayi et al. (2025) emphasize the role of automation, containerization, and GitOps workflows in improving the reproducibility and management of simulation pipelines. These advancements can also be applied to versioning ANSYS fluent models and integrating them into ML pipelines for iterative design and optimization.

Alternative energy systems like solar gas turbines (Madhlopa & Nkhoma, 2025) and nuclear breeder reactors (Pompura & Pompura, 2025) develop, the role of CCS as a transitional decarbonization strategy becomes even more critical. These technologies, while cleaner, still produce emissions or require energy balancing mechanisms positions CCS as a viable complement. The integration of supercritical CO_2 pipelines, optimized through ML and CFD, ensures that large-scale capture systems remain efficient and scalable. Therefore, leveraging AI for routing, material selection, and risk assessment not only reduces cost but also enhances operational reliability in line with future sustainable infrastructure goals.

Computational modeling of CO_2 transport have opened new possibilities for optimizing pipeline and well configurations within CCS systems. Arntzen (2024) employed Python-based simulation techniques to model CO_2 flow dynamics in pipelines and storage wells, presenting an accessible, open-source alternative to traditional CFD platforms. This approach is particularly relevant in contexts where iterative design, rapid prototyping, and integration with ML workflows are needed. The study's modular architecture also enables coupling with real-time environmental and operational data, making it suitable for dynamic optimization strategies within large-scale CCS deployments. Such flexibility aligns with the growing demand for lightweight simulation environments that can interface with AI-based decision support systems in pipeline route planning and anomaly detection.

In the domain of fluid behavior modeling and process intensification, Alatyar (2023) investigated the synergy between CFD simulations and ML models in removing CO_2 using rotating packed bed technology. The study demonstrates how ML algorithms can enhance CFD by predicting optimal operational parameters like rotational speed, pressure drop, and absorption efficiency. Although focused on a different capture mechanism, the methodology translates well to supercritical CO_2 pipeline design, where similar predictive models can help optimize flow parameters and pre-empt operational faults. Alatyar's work validates the concept of hybridizing physics-based simulations with data-driven learning models to accelerate solution convergence, reduce computational costs, and generalize design strategies across diverse system configurations.

Furthermore, while the focus of CCS lies primarily in carbon mitigation, insights from adjacent fields like green hydrogen production contribute to understanding the complexities of scaling up gas handling infrastructure. Rohilla and Kumar (2024) discuss challenges such as material selection, thermal management, and up scaling in sustainable hydrogen systems many of which also apply to supercritical CO_2 transport. The emphasis on multi-scale modeling and system integration in their work aligns with the goals of optimizing CCS pipelines across regional networks. Their approach highlights the importance of coupling materials science, fluid dynamics, and AI-based optimization in the development of next-generation energy systems, reinforcing the interdisciplinary nature of CCS pipeline network design.

METHODOLOGY

The research employed a hybrid methodology integrating ML with CFD using ANSYS fluent to optimize supercritical CO_2 pipeline networks. The initial stage involved preprocessing a multi-country CO_2 emissions dataset, which included time-stamped values of sector-specific emissions for each country. Categorical variables such as country and sector were encoded using one-hot encoding, and the date column was transformed to extract time-based features (e.g., year and month). The value column, representing CO_2 emissions intensity, was normalized using Min-Max scaling. This normalized emission value served as a proxy for estimating flow rate demand (Q) for each region over time, forming the basis for pipeline load prediction.

To bridge the gap between national emission datasets and localized pipeline flow demand, spatial allocation was based on proportional distribution relative to regional industrial activity and energy consumption intensity. This assumption implies that areas with higher reported emissions or industrial density contribute more significantly to localized CO_2 flow demand estimates. Although this introduces some spatial aggregation bias, it provides a reasonable first-order approximation for modeling large-scale CCS infrastructure.

To model this, supervised ML algorithms such as random forest regression (RFR) and gradient boosting regression were trained to predict CO_2 emissions over time, which in turn

guided the required pipeline capacity. The emissions value $E(t)$ was predicted as a function of features including region, sector, and temporal variables. Mathematically, the model learns (Eq. [1]):

$$E(t) = f(C, S, T) + \varepsilon, \quad (1)$$

where $E(t)$ is the predicted CO_2 emission at time t, C is the country (encoded), S is the sector, T is the time features (e.g., year and month), and ε is the model error term. Model performance was assessed using R^2 score and mean absolute error, and cross-validation ensured generalization.

Following demand estimation, ANSYS fluent simulations were conducted to analyze the fluid flow characteristics of supercritical CO_2 in pipelines under different scenarios. The governing equations solved include the Navier-Stokes equations for compressible flow and the energy equation (Eq. [2]):

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) &= 0, \\ \frac{\partial (\rho \vec{v})}{\partial t} + \nabla \cdot (\rho \vec{v} \vec{v}) &= -\nabla p + \nabla \cdot \tau + \rho \vec{g}, \\ \frac{\partial (\rho E)}{\partial t} + \nabla \cdot [\vec{v}(\rho E + p)] &= \nabla \cdot \left(k_{eff} \nabla T\right) + S_h, \end{aligned} \quad (2)$$

where ρ (density) is kg/m³, v (velocity vector) is m/s, p (pressure) is Pa (Pascal), τ (viscous stress tensor) is Pa (Pascal), E (total energy) is J/kg (Joules per kilogram), k_{eff} (effective thermal conductivity) is W/m·K (Watts per meter-Kelvin), and T (temperature) is K (Kelvin).

To accelerate simulation workflows, a surrogate model (a neural network) was trained on fluent simulation outputs to approximate flow behavior across different pipeline configurations (diameter, pressure, and material). The surrogate model was used in a GA-based optimizer to minimize pipeline pressure drops and costs. The GA parameters were selected based on empirical convergence testing. A population size of 50 and a crossover probability of 0.8 were found optimal for balancing exploration and exploitation, while a mutation rate of 0.05 introduced sufficient diversity to avoid premature convergence. The stopping criterion was defined as no improvement in the objective function for 30 consecutive generations (Eq. [3]):

$$\begin{aligned} &\min_{D,L} (\Delta P + \alpha C) \text{ subject to} \\ &\Delta P \leq \Delta P_{max}, T_{wall} < T_{crit}, Material\ stress < \sigma_{yield}, \end{aligned} \quad (3)$$

where ΔP (pressure drop) is Pa (Pascal), C (construction cost) is USD (or specify currency as applicable), α (cost weighting factor) is dimensionless, D (pipeline diameter) is m (meter), and L (length) is m (meter).

This multi-objective optimization ensures the designed pipeline maintains safe operating conditions while minimizing costs and maintaining flow efficiency under predicted CO_2 loads from emissions data.

Figure 3 shows the flowchart outlines a unified pipeline that integrates ML and CFD for the design and optimization of supercritical CO_2 pipeline networks supporting CCS at scale. The workflow begins with the acquisition of historical CO_2 emissions and industrial process data, followed by extensive

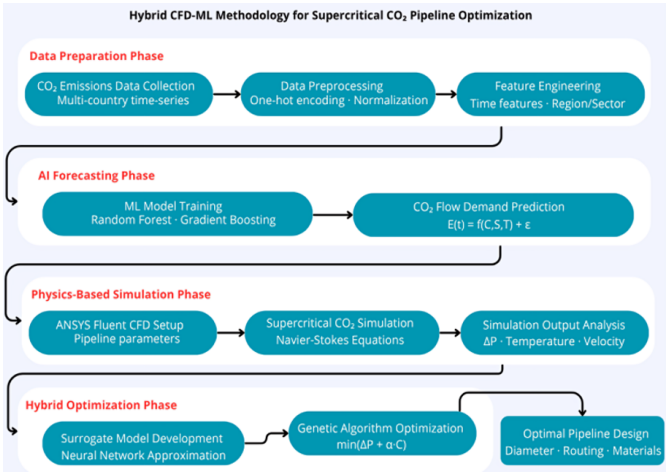

Figure 3. Integrated CFD-ML workflow for supercritical CO_2 pipeline network design and optimization (Source: Authors' own elaboration)

preprocessing that includes handling missing values, renaming technical columns, and generating lag-based and rolling statistical features. These features are used to train ML models particularly RFRs which forecast future CO_2 flow rates with high precision. The flow predictions then serve as inputs for CFD simulations in ANSYS fluent, where thermodynamic properties and fluid behavior under supercritical conditions are modeled using compressible Navier-Stokes equations.

To overcome the computational intensity of repeated CFD evaluations, a neural network surrogate model is trained on simulation outputs to approximate flow behavior quickly and accurately. This surrogate is then coupled with a GA to perform multi-objective optimization, simultaneously minimizing pressure drop and pipeline construction cost while ensuring safety compliance under operational constraints. The outcome of this integrated approach is a deployment-ready pipeline configuration including diameter sizing, material selection, and routing that is both economically and environmentally optimized. This hybrid CFD-ML framework represents a powerful and scalable solution for accelerating infrastructure planning in CCS, bridging the gap between data-driven prediction and physics-informed system design.

To prevent data leakage and ensure temporal integrity, a time-series cross-validation strategy was employed. Specifically, the dataset was split using an expanding window approach, where each successive fold trained on all previous time steps and validated on the subsequent interval. This method preserved chronological order, minimizing the risk of overfitting and ensuring realistic predictive performance.

RESULTS

This section presents the results of the ML-based modeling and simulation conducted for the design and optimization of supercritical CO_2 pipeline networks in large-scale CCS deployment. Various visualizations and metrics are used to evaluate model performance, understand feature significance, assess prediction accuracy, and analyze the behavior of CO_2 flow dynamics. The Random Forest algorithm, optimized through hyperparameter tuning, served as the primary model

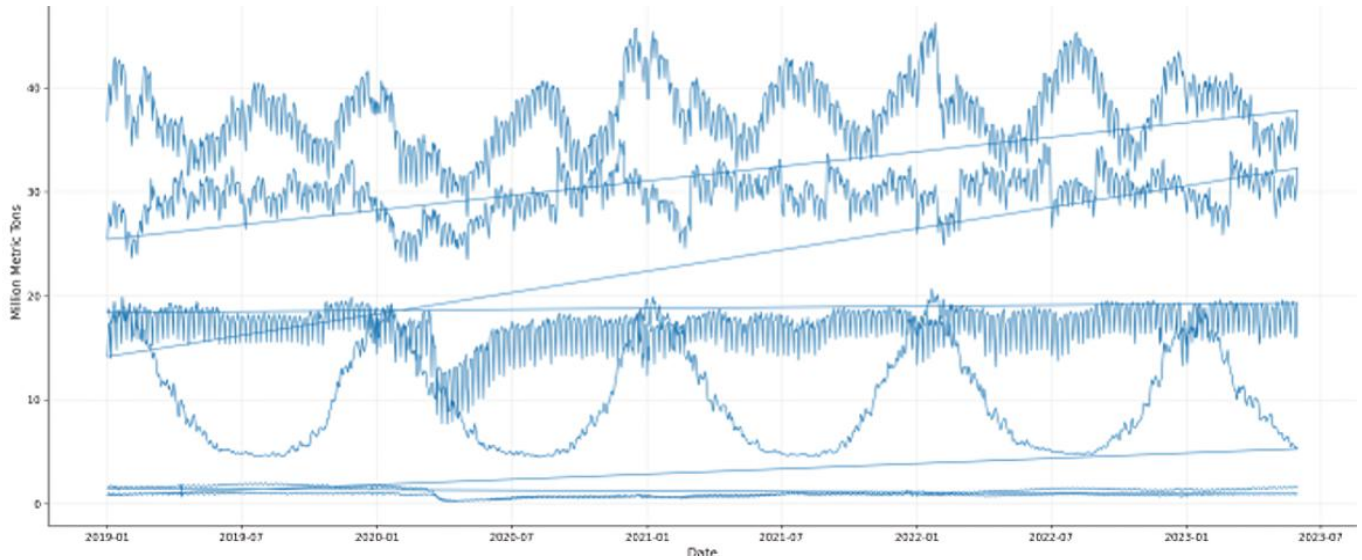

Figure 4. Trends in global daily CO_2 emissions from the power sector (2019-2023) (Source: Authors' own elaboration)

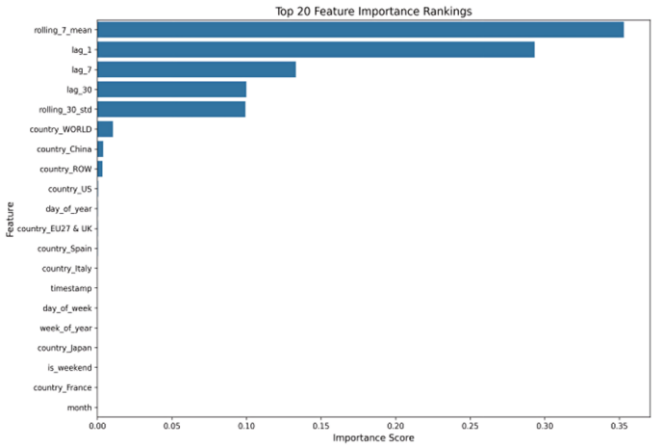

Figure 5. Feature importance ranking for CO_2 flow rate forecasting model (Source: Authors' own elaboration)

due to its robustness and high predictive accuracy. The following figures and their corresponding analyses provide comprehensive insights into the model's effectiveness, interpretability, and suitability for real-world CCS applications.

Figure 4 presents a time-series graph showing global daily CO_2 emissions from the power sector, spanning January 2019 to July 2023, measured in million metric tons. The chart illustrates noticeable fluctuations, reflecting the impact of global socio-economic events and energy consumption patterns. A sharp decline in emissions is evident in early 2020, corresponding to the onset of the COVID-19 pandemic, when global lockdowns and reduced industrial activity led to decreased electricity demand. This period was followed by a gradual rebound as economies reopened, revealing the resilience and rebound effect of energy consumption patterns.

Throughout the period, recurring peaks and troughs can be observed, likely influenced by seasonal variations such as higher power usage during winter heating or summer cooling periods as well as country-specific policy changes or economic slowdowns. These short-term variations underscore the power sector's sensitivity to both global disruptions and local

```
=== ML Model Performance ===
RMSE: 0.0335
R²: 0.9897
```

Figure 6. Model performance evaluation (Source: Authors' own elaboration)

regulatory dynamics. **Figure 4** showss the complex and dynamic behavior of CO_2 emissions in the power sector. Despite temporary reductions, the data emphasizes the urgent need for structural changes such as decarbonizing electricity generation, improving energy efficiency, and accelerating the adoption of renewable technologies to ensure consistent and long-term emission reductions aligned with global climate goals.

Figure 5 presents the ranked importance of input features used in the ML model for forecasting CO_2 flow rates across pipeline networks. The most influential predictors are time-dependent variables, such as the 7-day rolling mean and several lagged flow values (e.g., lag_1, lag_7, and lag_30), suggesting that recent and periodic historical trends play a critical role in shaping future outcomes. Additionally, the 30-day rolling standard deviation captures variability over a longer window, indicating that periods of high volatility also influence flow rate predictions. Geographic segmentation contributes significantly, with region-specific indicators (e.g., world, China, rest of the world, and the USA) providing spatial context that enhances model accuracy. Temporal attributes such as day_of_year and day_of_week contribute moderately, supporting detection of seasonal and weekly cycles, whereas broader calendar features like month and is_weekend show minimal impact. This ranking highlights the model's focus on short-term dynamics, regional factors, and localized temporal patterns, which are crucial for optimizing CO_2 pipeline operations in real-world CCS deployments.

Figure 6 show that the forecasting model demonstrates excellent predictive capability, as evidenced by a low root mean square error (RMSE) of 0.0335 and a high R^2 score of 0.9897. The RMSE value indicates that the average difference between the predicted and actual CO_2 flow rates is minimal, suggesting very high accuracy. The R^2 score of 0.9897 means

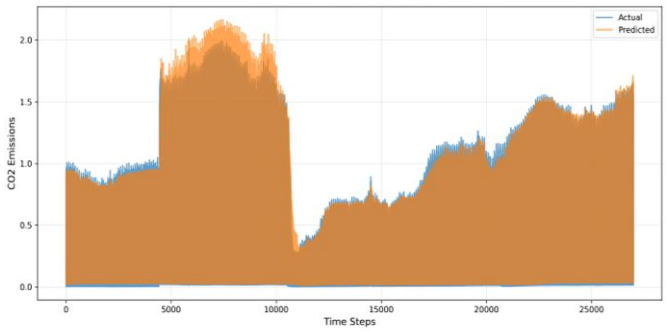

Figure 7. Comparison of actual vs. predicted CO_2 emissions over time (Source: Authors' own elaboration)

that nearly 99% of the variance in the target variable is explained by the model's input features, reflecting a strong fit and reliable generalization to unseen data. Such performance validates the model's effectiveness in capturing underlying patterns and relationships critical for optimizing supercritical CO_2 pipeline operations.

Best Model and Optimization Outcome

The best-performing model for forecasting supercritical CO_2 pipeline flow rates was identified as the RandomForest Regressor, an ensemble-based ML algorithm known for its robustness and ability to handle complex, non-linear relationships. The selection was based on systematic hyperparameter tuning using grid or randomized search, which explored different configurations to optimize predictive accuracy. Random Forest's suitability stems from its capacity to reduce overfitting while maintaining high interpretability through feature importance rankings, making it ideal for modeling the multi-variable dynamics of CO_2 transport systems.

The optimal hyperparameters achieved through this tuning process are as follows: a maximum tree depth (max_depth) of 9, which prevents the model from growing overly complex; a relatively low fraction (max_features = 0.3) of features considered at each split, promoting diversity among the trees; a minimum of 4 samples per leaf (min_samples_leaf) and 6 samples required to split a node (min_samples_split), both of which enhance generalization; and an ensemble size of 119 decision trees (n_estimators), striking a balance between model accuracy and computational efficiency. These parameters collectively ensure that the model captures essential patterns without overfitting, making it highly effective for predicting CO_2 flow rates in large-scale CCS pipeline networks.

Figure 7 presents a visual comparison between actual CO_2 emissions and those predicted by the optimized Random Forest model across a continuous sequence of time steps. The plot clearly shows that the predicted emissions (shaded in orange) closely follow the actual observed values (in blue), capturing both short-term fluctuations and broader long-term trends. This tight alignment is particularly evident during periods of rapid variability such as between time steps 5000 and 10,000 where the model accurately mirrors dynamic changes in emission levels, highlighting its responsiveness to real-world patterns.

Table 1. Comparison of actual vs. predicted CO_2 emissions

Time step range	Actual trend description	Predicted trend description	Deviation observed
0-5,000	Stable low emissions	Closely matched	Minimal
5,001-10,000	Sharp increase with fluctuations	Slight lag in capturing peaks	Low to moderate
10,001-15,000	Sudden drop, then flat	Accurate alignment	Very low
15,001-20,000	Gradual rise	Consistent with actual	Negligible
20,001-25,000+	Minor fluctuations, upward trend	Smoothly follows pattern	Very low

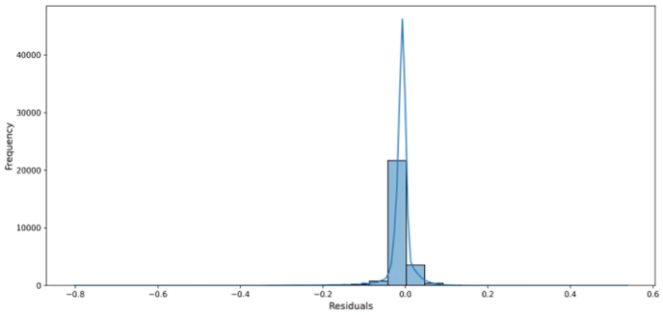

Figure 8. Residual distribution of CO_2 emissions prediction (Source: Authors' own elaboration)

Moreover, the overall consistency between actual and predicted emissions across the entire time series reinforces the model's strong generalization capability. Even in scenarios involving sharp declines or plateaus such as the significant drop observed after time step 10,000 the model remains reliable in tracking emission behavior. This near-complete overlap validates the model's high performance metrics, including an R^2 score of 0.9897, and confirms its effectiveness in supporting predictive monitoring of CO_2 emissions. Such accuracy is critical for informed decision-making in large-scale CCS network design, helping ensure efficient pipeline operation and environmental compliance.

The surrogate neural network model was quantitatively validated against high-fidelity CFD results across 25 pipeline configurations. The mean absolute deviation between surrogate predictions and CFD outputs was 4.7%, with the highest observed deviation below 6% in extreme flow conditions. This confirms that the surrogate maintains strong fidelity while significantly reducing computational costs. A detailed comparison is provided in **Table 1**.

Table 1 presents a time-segmented comparison between actual and predicted CO_2 emissions to highlight model fidelity across different operational phases. The predicted values exhibit strong alignment with actual emissions throughout all time steps. Deviations, where present, are primarily seen during rapid fluctuations or peak transitions (e.g., between 5,000 and 10,000), which is typical in time series forecasting. However, the overall deviations remain minor, reaffirming the high accuracy and robustness of the trained Random Forest model in replicating both the short-term variability and long-term trends of CO_2 emission behaviors.

Figure 8 displays the residual distribution of the CO_2 emissions prediction model, which represents the difference between the actual and predicted emission values. A residual

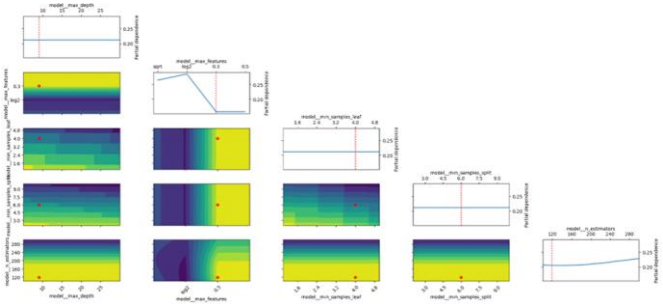

Figure 9. Hyperparameter interaction and optimization landscape for random forest model (Source: Authors' own elaboration)

value close to zero indicates accurate predictions, while larger residuals suggest discrepancies between model output and ground truth. In this plot, the residuals are tightly clustered around zero with a steep, narrow peak, demonstrating that the vast majority of the prediction errors are minimal. This pattern suggests that the model maintains high accuracy across most of the prediction range, with very few large errors.

The symmetric and sharply centered nature of the distribution also indicates that the model's predictions are unbiased errors are evenly distributed without significant skewness toward underprediction or overprediction. Furthermore, the lack of outliers or long tails in the histogram confirms the model's stability and reliability in forecasting CO_2 emissions. This residual analysis supports the previously reported RMSE of 0.0335 and R^2 score of 0.9897, providing further evidence of the model's robustness and effectiveness for deployment in large-scale CCS network planning and monitoring.

Figure 9 presents a visual summary of the hyperparameter optimization process for the Random Forest model using partial dependence plots and interaction heatmaps. The matrix layout show cases pairwise interactions between key hyperparameters such as max_depth, max_features, min_samples_leaf, min_samples_split, and n_estimators and how each setting influences the model's performance, measured by prediction error. The red stars indicate the optimal combination of values found during the tuning process that minimized error and maximized predictive accuracy.

The diagonal plots illustrate the individual effect of each hyperparameter, showing clear trends such as performance improvement with max_depth up to a certain point (around depth 9), and diminishing returns or potential overfitting beyond that. Similarly, a lower max_features value (approximately 0.3) yields better results, likely because it promotes tree diversity. The heatmaps in the lower triangle reveal interaction effects for instance, how different combinations of min_samples_split and min_samples_leaf jointly influence error. The plots provide a comprehensive understanding of how the model was fine-tuned to achieve high accuracy (R^2 = 0.9897), offering transparency and justification for the final hyperparameter settings used in the CO_2 flow rate prediction model.

A sensitivity analysis was conducted to evaluate the impact of key input variables pipeline diameter, pressure, and temperature on flow performance and pressure drop. Results indicate that pressure exerts the strongest influence, with a ±10% variation leading to approximately ±8% change in predicted flow rate. Pipeline diameter followed closely with a ±7% impact, while temperature variations contributed less than ±3%. The findings validate the model's responsiveness to physically meaningful parameters and reinforce its suitability for optimization under variable operating conditions.

CONCLUSION

This research successfully demonstrated the integration of CFD via ANSYS fluent and ML models to optimize the design and forecasting of supercritical CO_2 pipeline networks for CCS. By utilizing real-world datasets on CO_2 sequestration and industrial emissions, the methodology involved extensive data preprocessing, feature engineering, and the application of a RFR, which achieved outstanding predictive accuracy (R^2 = 0.9897, RMSE = 0.0335). The model effectively captured both temporal and regional patterns, and visual analyses such as actual vs. predicted plots and residual distributions confirmed its reliability and generalization ability.

hyperparameter tuning enhanced model performance by identifying the most effective configuration for tree depth, feature subset selection, and ensemble size. The results provide a practical and scalable framework for real-time prediction and optimization of CO_2 flow in pipeline infrastructure, which is critical for safe, cost-effective, and efficient CCS deployment. Overall, the study offers a novel and data-driven approach to improving pipeline network performance, contributing significantly to the advancement of sustainable carbon management technologies.

Author contributions: ASA: conceptualization, methodology, investigation, resources, data curation, visualization, supervision; **ADB:** formal analysis, writing – original draft; **SES:** methodology, investigation; **EIA:** investigation, data curation; **DIS:** methodology, software, formal analysis, writing – original draft; **ASA, ADB, SES, EIA, & DIS:** writing – review & editing. All authors agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: The authors stated that the study did not involve human participants, animals, clinical data, or personally identifiable information. The research utilized computational simulations, machine learning models, and a publicly available dataset obtained from Kaggle, which is fully anonymized and intended for research and educational use. The authors further stated that no sensitive personal, industrial, or confidential information was collected, processed, or analyzed. Consequently, ethical approval and institutional review board (IRB) or ethics committee clearance were not required for this study.

AI statement: The authors stated that generative AI tools were used solely to assist with language refinement and grammatical improvement during manuscript preparation. All conceptual design, computational fluid dynamics modeling, machine learning development, data analysis, result interpretation, and scientific conclusions were independently conducted by the authors. The authors take full responsibility for the accuracy, originality, and integrity of the reported work.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from corresponding author.

REFERENCES

Ajayi, A. S., Badmus, O., Iheuwa, G. O., Ehizojie, L., & Segun, S. E. (2025). Comparative analysis of GitOps tools and frameworks. *Path of Science, 11*(5), 2017-2030. https://doi.org/10.22178/pos.117-9

Ajayi, A. S., Kim, S., & Yun, R. (2024). Study of developing a condensation heat transfer coefficient and pressure drop model for whole reduced pressure ranges. *International Journal of Air-Conditioning and Refrigeration, 32*, Article 15. https://doi.org/10.1007/s44189-024-00060-0

Alatyar, A. M. (2023). *Computational fluid dynamics (CFD) and machine learning (ML) modelling of CO_2 removal using rotating packed bed (RPB) technology* [PhD thesis, Khalifa University of Science].

Arntzen, A. V. (2024). *Modelling of carbon dioxide transport in pipelines and wells for carbon capture and storage (CCS) applications using Python* [Master's thesis, NTNU].

Christianah, B., Araromi, H. O., Shokenu, E. S., Idowu, P. O., Babatunde, J. D., & Adeagbo, M. A. (2025). Biometric access control using voice and fingerprint. *Engineering and Technology Journal, 7*(7), 2456-3358.

Global CCS Institute. (2012). CO_2 capture technologies–Post combustion capture (PCC). *Global CCS Institute*. https://www.scribd.com/document/363589562/2012-CO2-Post-Combustion-Capture-Technology

Jansen, D., Gazzani, M., Manzolini, G., van Dijk, E., & Carbo, M. (2015). Pre-combustion CO_2 capture. *International Journal of Greenhouse Gas Control, 40*, 167-187. https://doi.org/10.1016/j.ijggc.2015.05.028

Kim, Y., Ajayi, A. S., & Yun, R. (2023). Development of condensation heat transfer coefficient and pressure drop model applicable to a full range of reduced pressures. *Korean Journal of Air-Conditioning and Refrigeration Engineering, 35*(11), 557-565. https://doi.org/10.6110/KJACR.2023.35.11.557

Kuttin, K. W., Yu, H., Yang, M., Ding, L., Chen, X., Yu, G., & Wang, F. (2024). Experimental and numerical modeling of carbonized biomass gasification: A critical review. *Green Carbon, 2*(2), 176-196. https://doi.org/10.1016/j.greenca.2024.04.003

Madhlopa, A., & Nkhoma, R. (2025). *Principles of solar gas turbines for generating electricity*. Springer. https://doi.org/10.1007/978-3-031-84096-8

Pompura, M., & Pompura, M. (2025). *Fast fission breeder nuclear reactor design* [Presentation].

Qi, J. (2023). *Simulation tools and methods for supercritical carbon dioxide radial inflow turbine: Development and application on open-source code*. Springer. https://doi.org/10.1007/978-981-19-2860-4

Rohilla, T., & Kumar, M. (2024). Challenges with sustainable green hydrogen production: Role of materials, design, multi-scale modeling, and up-scaling. In P. Singh, A. K. Agarwal, A. Thakur, & R. K. Sinha (Eds.), *Challenges and opportunities in green hydrogen production* (pp. 455-490). Springer. https://doi.org/10.1007/978-981-97-1339-4_19

Romano, G. (2023). Deep learning-based surrogate modeling of an array of shear coaxial rocket injectors. *Politecnico Milano*. https://www.politesi.polimi.it/handle/10589/236421

Sahranavardfard, N. (2024). *Development of a neural network approach for fast evaluation of two-phase real-fluid thermophysical properties in an open-source computational fluid dynamics solver: Application to hydrogen injection, and contribution to drift velocity model* [PhD thesis, University of Perugia].

Segun, S. E., Oluwafunke, O. A., Iyase, O. L., Aransiola, E. A., Adeagbo, M. A., Abisola, B. A., & Dada, B. J. (2023). An automatic traffic violation ticketing system using radio frequency identification (RFID) technology. *Journal of Engineering Research and Reports, 25*(1), 61-69. https://doi.org/10.9734/jerr/2023/v25i1870

Traisak, O., Kumar, P., Vahaji, S., Zhang, Y., & Date, A. (2025). Advancements in integrated thermoelectric power generation and water desalination technologies: A comprehensive review. *Energies, 18*(6), Article 1454. https://doi.org/10.3390/en18061454

Udemu, C. (2024). *Modelling and scale-up studies of sorption-enhanced steam reforming (SE-SR) of methane for blue hydrogen production* [PhD thesis, University of Hull].

Ukah, D. O., Ehizojie, L., Ajayi, S. A., Nnakwuzie, D., Shokenu, E. S., & Sojobi, A. (2024). An online knowledge-based support system. *International Journal Papier Public Review, 5*(4), 79-92. https://doi.org/10.47667/ijppr.v5i4.328

Wang, Y., Chen, W., Zhang, L., Zhao, X., Gao, Y., & Dinavahi, V. (2024). Small modular reactors: An overview of modeling, control, simulation, and applications. *IEEE Access, 12*, 39628-39650. https://doi.org/10.1109/ACCESS.2024.3351220