Earth Science: Environmental Challenges and Solutions

Research Article

Open Access Full Text Article

Application of Machine Learning to Enhanced Prediction and Optimization of CO_2 and CH_4 Emission Reduction Potential

Ajayi Abiola Samuel[1*], Shokenu Emmanuel Segun[2], Ajayi Eniola Isaac[3], Godwin Iheuwa[4], Ayoola David Bodude[5], Brian Cobb Hanrahan[6], Isaac Njoku[7]

[1]Department of Computer Science, Texas Southern University, USA.
[2]Department of System Engineering, African University of Science and Technology, Nigeria.
[3]Aptech Computer Education, Aptech Isheri-Berger center, Ojodu Berger, Lagos, Nigeria.
[4]Department of Computer Science, University of Bedfordshire, UK.
[5]Department of Chemical Engineering, University of Strathclyde, Galsgow, UK.
[6]World Aerospace, Discovery Business Park, Cedar Park, Texas, USA.
[7]Department of Chemical Engineering, Colorado State University, USA.

Correspondence:

Ajayi Abiola Samuel
Department of Computer Science, Texas Southern University, USA.

Received: November 27, 2025; **Published:** December 16, 2025

How to cite this article:

Ajayi, A. S., Shokenu, E. S., Ajayi, E. I., Iheuwa, G., Ayoola, D. B., Hanrahan, B. C., & Isaac, N. (2025) 'Application of machine learning to enhanced prediction and optimization of CO_2 and CH_4 emission reduction potential', *Earth Science: Environmental Challenges and Solutions*, vol. 1, no. 1, pp. 1–7.

Abstract

This study presents a machine learning-based framework for enhancing the prediction and optimization of CO_2 and CH_4 emission reduction potential using multi-sectoral and socio-economic data, aligned with Sustainable Development and climate action goals. Leveraging Random Forest Regression, the model achieved exceptional predictive performance ($R^2 \approx 0.997$, RMSE ≈ 53.69), with predicted emissions closely matching observed values and minimal systematic bias. Feature importance analysis identified oil production, coal-related emissions, and other CO_2 sources as the dominant contributors, while GDP and cement production exhibited moderate influence. Correlation analysis revealed strong interdependencies between greenhouse gas emissions and factors such as population, N_2O emissions, and fossil fuel consumption, underscoring the interconnected nature of emission drivers. The novelty of this approach lies in integrating high-resolution data with advanced predictive modeling to not only forecast emissions accurately but also pinpoint priority areas for targeted mitigation strategies. The findings provide a scalable, evidence-based decision-support tool for policymakers, enabling them to design effective interventions that accelerate decarbonization, methane reduction, and broader Sustainable Development objectives.

Keywords: Machine Learning, CO_2 Emissions, CH_4 Emissions, Feature Importance, Predictive Modeling

Introduction

Carbon dioxide (CO_2) and methane (CH_4) are among the most impactful greenhouse gases (GHGs) influencing global climate change due to their strong radiative forcing and significant contribution to global warming (Filonchyk et al. 2024; Pulles and Van Amstel 2010). While CO_2 remains the most abundant anthropogenic GHG, primarily from fossil fuel combustion and deforestation (Brander and Davis 2012), CH_4 is particularly concerning because of its high global warming potential over 80 times that of CO_2 on a 20 years timescale despite its lower atmospheric concentration (Pulles and Van Amstel 2010). Studies highlight that both gases drive increases in average global temperature, disrupt precipitation patterns, and intensify extreme weather events, posing threats to biodiversity, food security, and human well-being (Filonchyk et al. 2024; Faraday and Oluwabunmi 2024). The persistence and potency of these gases make their reduction a global priority.

To address these challenges, international frameworks such as the Paris Climate Agreement set ambitious targets to limit temperature rise to well below 2 °C, with efforts to stay within 1.5 °C above pre-industrial levels (Li et al. 2024; Kreibich 2024). Meeting these goals requires significant reductions in both CO_2 and CH_4, with CH_4 control offering faster climate benefits due to its shorter atmospheric lifespan (Pulles and Van Amstel 2010). However, as Alagade and Sahu (2025) note in their satellite-based greenhouse gas forecasting work, current emission trajectories suggest that existing measures are insufficient to meet climate goals. This underscores the need for data-driven, predictive, and optimization-based strategies that can identify high-impact interventions and accelerate emission reduction progress.

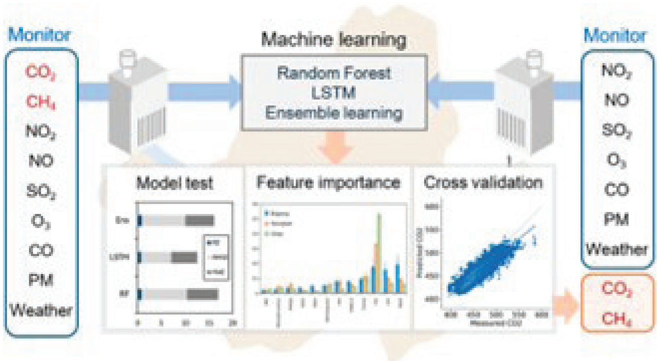

Figure 1. *Machine learning framework for predicting CO_2 and CH_4 concentrations. Adapted from Park et al. (2025).*

Figure 1 shows a machine learning based framework for predicting CO_2 and CH_4 emissions using environmental and atmospheric monitoring data, where inputs such as NO_2, NO, SO_2, O_3, CO, particulate matter (PM), and weather parameters are processed through models like Random Forest, LSTM, and ensemble learning to generate predictive insights on greenhouse gas concentrations. The framework emphasizes model evaluation through testing, feature importance ranking, and cross-validation, ensuring reliable capture of emission dynamics. Beyond prediction, such approaches can be integrated with technological solutions, including smart exhaust after-treatment systems and catalytic converters that dynamically optimize emission control (Wu et al. 2024; Bakhchin et al. 2024), and with zeolite-based adsorbents for efficient carbon capture from point sources (Zhang et al. 2025). They also align with advances in materials science, such as molecular dynamics simulations for CO_2 and CH_4 hydrate replacement in porous media to support methane recovery and carbon sequestration (Zhang et al. 2024), as well as nature-based strategies like enhancing plant diversity to improve soil carbon storage (Dang et al. 2024). Together, this highlights the potential of machine learning as a predictive backbone that strengthens both technological and ecological pathways for greenhouse gas reduction.

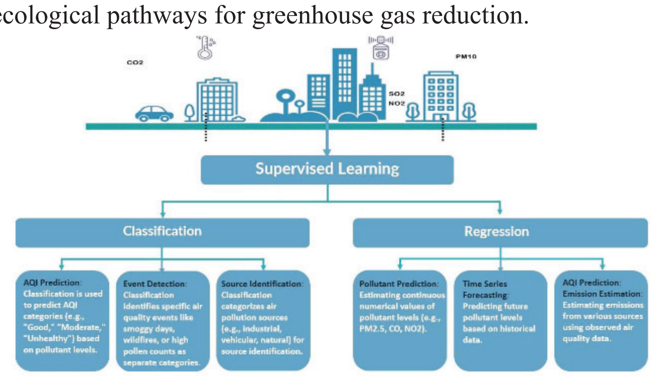

Figure 2. *Supervised learning framework for air quality prediction. Reprinted from Essamlali et al. (2024).*

Figure 2 illustrate a supervised machine learning diagram that offers valuable tools for urban air quality prediction by using monitored data such as CO_2, SO_2, NO_2, and PM_{10}. Through classification, models can predict AQI categories, detect pollution events, and identify emission sources, while regression approaches estimate pollutant levels, forecast future air quality, and quantify emissions (Essamlali et al. 2024). Beyond monitoring, machine learning complements technological solutions like smart exhaust after-treatment systems, catalytic converters, and carbon capture and utilization technologies by enhancing predictive accuracy and optimization of emission control (Wu et al. 2024; Bakhchin et al. 2024). By integrating diverse datasets, including socio-economic indicators, industrial activity, and energy use, models such as gradient boosting, random forests, and neural networks can forecast emissions, identify key drivers, and assess the impacts of policy or technological interventions (Alagade and Sahu 2025; Dang et al. 2024). These predictive insights support innovations such as zeolite-based adsorbents for CO_2 capture (Zhang et al. 2025), hydrate replacement for methane recovery and sequestration (Zhang et al. 2024), and nature-based solutions like enhancing plant diversity to boost soil carbon storage (Dang et al. 2024). Ultimately, optimization algorithms applied to ML outputs can identify efficient intervention strategies for reducing CO_2 and CH_4 while aligning with global climate goals.

Despite the availability of emissions data, advanced statistical models, and mitigation technologies, there is still a limited integration of machine learning-based prediction with optimization frameworks that jointly address CO_2 and CH_4 reduction potential. This research seeks to fill that gap by developing a machine learning-enhanced framework to analyze patterns and drivers of CO_2 and CH_4 emissions, forecast future emission trends under various intervention scenarios, and optimize strategies to maximize reduction potential. In doing so, it builds upon prior findings in greenhouse gas science (Filonchyk et al. 2024; Pulles and Van Amstel 2010; Brander and Davis 2012), technological innovation (Wu et al. 2024; Bakhchin et al. 2024; Zhang et al. 2025; Zhang et al. 2024), and predictive analytics (Alagade and Sahu 2025; Dang et al. 2024), while aligning with international climate policy objectives (Li et al. 2024; Kreibich 2024).

Related Works

Global CO_2 and CH_4 emissions have been extensively studied due to their significant impact on climate change and environmental sustainability. CO_2 emissions primarily result from fossil fuel combustion, industrial processes, and deforestation, while CH_4 emissions are largely driven by agriculture, waste management, and energy production (Pulles and Van Amstel 2010; Brander and Davis 2012). Filonchyk et al. (2024) highlight that the combined radiative forcing of these gases accelerates global warming, with CH_4 being over 80 times more potent than CO_2 on a 20-year horizon. Recent trends indicate a continuous rise in global emissions, despite international agreements such as the Paris Climate Accord, underscoring the urgency for advanced monitoring and mitigation approaches (Li et al. 2024; Kreibich 2024).

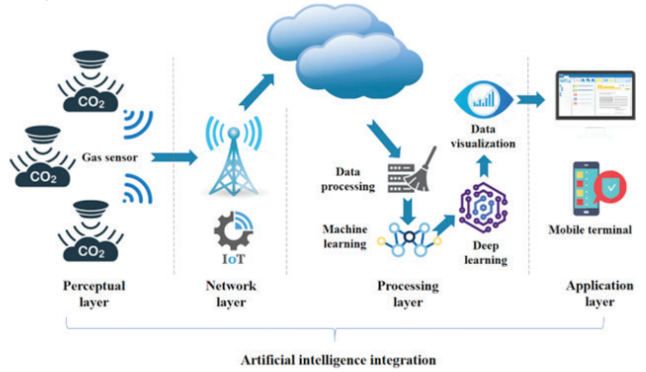

Figure 3. *Framework integrating AI and IoT for real-time monitoring and prediction of building carbon emissions. Adapted from Hua et al. (2025).*

Figure 3 show a Machine learning (ML) combined with IoT provides powerful tools for predicting and optimizing carbon emissions by integrating data from gas sensors, industrial activity, energy use, and environmental monitoring (Alagade and Sahu 2025; Dang et al. 2024). Advanced models such as gradient boosting, random forests, and neural networks can forecast emissions, identify key drivers, and evaluate the effectiveness of policy or technological interventions (Wu et al. 2024; Bakhchin et al. 2024). As illustrated in AI–IoT frameworks, these approaches enable real-time monitoring and decision support, ensuring efficient CO_2 and CH_4 mitigation aligned with global climate goals (Zhang et al. 2024; Zhang et al. 2025; Hua et al. 2025).

Various strategies have been implemented to reduce greenhouse gas emissions, ranging from technological interventions to policy-driven frameworks. Technological solutions include catalytic converters, carbon capture and storage (CCS), zeolite-based CO_2 adsorption, and smart exhaust after-treatment systems capable of adapting in real-time to operating conditions (Wu et al. 2024; Bakhchin et al. 2024; Zhang et al. 2025). Innovative approaches, such as CO_2–CH_4 hydrate replacement in porous media for combined methane recovery and carbon sequestration, further broaden the mitigation toolkit (Zhang et al. 2024). On the policy front, global and regional agreements, coupled with sector-specific emission standards, aim to accelerate decarbonization while supporting sustainable economic growth (Li et al. 2024). However, the implementation of these strategies often faces challenges related to cost, scalability, and integration across multiple sectors.

In recent years, machine learning (ML) has emerged as a powerful tool for environmental data analysis, emissions forecasting, and optimization of mitigation strategies. ML models such as random forests, gradient boosting, and neural networks have been successfully applied to predict GHG emissions, identify key emission drivers, and evaluate intervention scenarios (Alagade and Sahu 2025; Dang et al. 2024). These methods outperform traditional statistical approaches in handling large, heterogeneous datasets and capturing complex non-linear relationships. Despite these advances, most existing studies focus on either CO_2 or CH_4 independently, target specific geographic regions, or employ predictive models without optimization capabilities. This creates a research gap for integrated frameworks that simultaneously address both CO_2 and CH_4, leveraging ML for joint prediction and optimization to inform cross-sectoral emission reduction strategies.

Machine learning (ML) has emerged as a versatile tool for enhancing greenhouse gas (GHG) reduction strategies by improving prediction accuracy and optimizing interventions across multiple sectors. In carbon capture, utilization, storage, and transportation, ML models enable precise efficiency assessments and risk prediction (Du et al. 2025), while global forecasting studies demonstrate their value in projecting reduction trajectories under energy transition scenarios (Gan and Zhao 2024). Applications extend to biomedical outcomes of CO_2 reduction (Shafaghat 2025), optimization of wastewater treatment within international guidelines (Kothale and Sadgir 2025), and predictive analytics for sustainable industrial production (Ojadi et al. 2023). Advanced hybrid neural networks have been developed to improve methane separation, CO_2 sequestration, and methane recovery from coal seams (Xue et al. 2024a; Xue et al. 2024b), while optimization techniques also support hydrogen production from biomass-derived methane (Ehinmowo et al. 2025). On the materials front, ML accelerates the discovery of covalent organic frameworks for energy and environmental applications (Wang et al. 2025) and provides quantitative insights into regional carbon neutrality policy synergies through deep learning (Zhang and Feng 2024). Collectively, these studies demonstrate that ML not only strengthens predictive modeling of CO_2 and CH_4 reduction but also optimizes sectoral interventions, thereby supporting global carbon neutrality and sustainability targets.

Methodology

This study adopts a data-driven machine learning framework to predict and analyze CO_2 and CH_4 emissions across multiple countries using historical environmental and socio-economic data. The methodology is structured into five stages: data acquisition, preprocessing, exploratory analysis, model development, and scenario simulation. The dataset, obtained from global_emissions.csv, contains annual emission values along with GDP, population, and sector-specific breakdowns such as coal, oil, gas, cement, flaring, and other sources. Preprocessing will include handling missing numerical values through mean imputation, creating composite indicators such as emissions per capita and emissions to GDP ratio, and standardizing features using Z-score normalization to prepare the data for model training.

Exploratory Data Analysis (EDA) will be conducted to identify patterns, assess emission trends, and examine relationships between socio-economic factors and emissions. Planned visualizations include correlation heatmaps to analyze inter-variable relationships and time series plots to illustrate historical trends for the top CO_2 emitting countries. For predictive modeling, Random Forest Regression will be implemented for both CO_2 and CH_4 due to its robustness in handling non-linear relationships and heterogeneous datasets. Hyperparameter tuning will be performed using grid search cross-validation to determine optimal settings such as tree depth, number of estimators, and minimum sample split size. Model performance will be evaluated using Root Mean Squared Error (RMSE) and the coefficient of determination (R^2), with cross-validation applied to enhance reliability and reduce overfitting risks.

To improve interpretability, feature importance analysis will be performed using both built-in Random Forest importance scores and permutation importance to identify the most influential socio-economic and sectoral factors. Scenario simulations will be used to assess the potential impact of policy or technological interventions; for example, modeling the effects of a 10% GDP increase combined with a 5% reduction in coal-related emissions to estimate projected changes in CO_2 output. The methodology also includes generating visual outputs such as predicted-versus-actual plots, feature importance charts, and emission time series to facilitate interpretation and support evidence-based climate policy planning. This structured approach is intended to deliver both predictive capability and actionable insights for emission reduction strategies.

Result

This section presents the results of the machine learning models developed to predict and analyze CO_2 and CH_4 emissions, highlighting their accuracy, feature importance, and predictive reliability. The findings provide key insights into the most influential emission sources, guiding targeted reduction strategies.

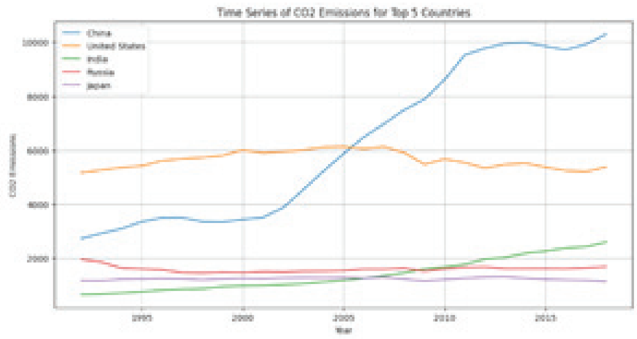

Figure 4: *Time Series Trends of CO_2 Emissions for the Top Five Emitting Countries*

Figure 4 show the CO_2 emission trends from 1990 to 2018 for the five highest emitting countries, highlighting critical patterns relevant to the application of machine learning in predicting and optimizing CO_2 and CH_4 emission reduction potential. China shows a sharp and sustained increase in emissions, surpassing all other nations after the early 2000s, largely due to rapid industrialization and coal dependent energy production. The United States maintains relatively high but stable emissions with a slight decline after 2007, reflecting economic shifts and the adoption of cleaner energy technologies. India's emissions display a steady upward trajectory, indicating growing industrial and energy demands, while Russia's emissions dropped significantly after the early 1990s economic transition and have since stabilized. Japan exhibits relatively flat emission levels with minor fluctuations. These distinct national trends underscore the importance of country-specific modeling and optimization strategies, as emission drivers and reduction opportunities vary significantly across economic, industrial, and energy contexts.

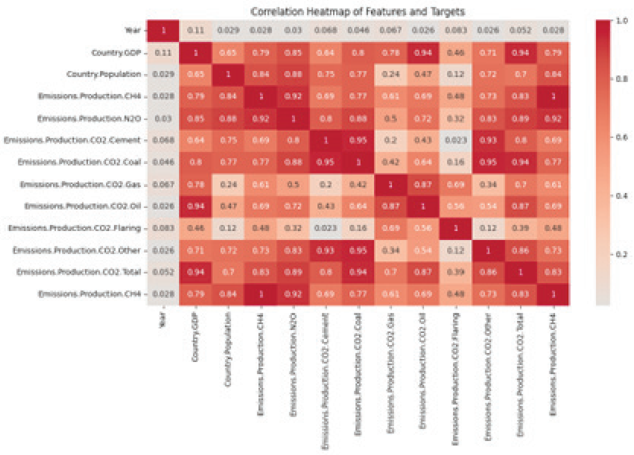

Figure 5: *Correlation Heatmap of Socio-Economic and Sectoral Factors with CO_2 and CH_4 Emissions*

Figure 5 shows the correlation heatmap in this study, focused on the application of machine learning to enhance the prediction and optimization of CO_2 and CH_4 emission reduction potential, reveals significant relationships between socio-economic indicators, sectoral emission sources, and greenhouse gas outputs. CO_2 total emissions exhibit strong positive correlations with GDP (0.94), coal related emissions (0.95), and oil-related emissions (0.94), indicating that economic growth and fossil fuel consumption are major drivers of CO_2 output.

Similarly, CH_4 emissions show high correlations with population (0.84), N_2O emissions (0.92), and other emission types, reflecting the interconnected nature of greenhouse gas emissions and demographic factors. These relationships suggest that changes in one emission source or socio-economic factor can influence multiple greenhouse gases simultaneously.

Understanding these correlations is critical for developing robust multivariate machine learning models capable of capturing such interdependencies. This insight enables the model to identify high-impact areas for intervention, thereby supporting the design of more effective, data-driven policies and technological measures aimed at reducing both CO_2 and CH_4 emissions.

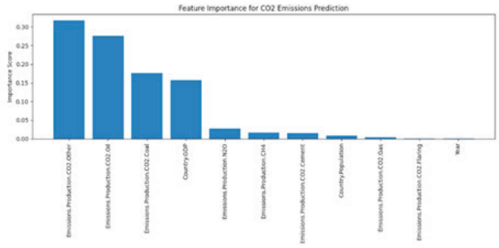

Figure 6: *Feature Importance Rankings for CO_2 Emissions Prediction using Random Forest Regression.*

Figure 6 show the feature importance analysis for CO_2 emissions prediction shows that Emissions.Production.CO_2.Other, Emissions. Production.CO_2.Oil, and Emissions.Production.CO_2.Coal are the most influential variables, together contributing over 75% of the model's predictive power. GDP and N_2O emissions also have moderate influence, while factors such as population, gas, and flaring emissions have minimal impact. This indicates that sector-specific CO_2 sources dominate the prediction of total CO_2 emissions.

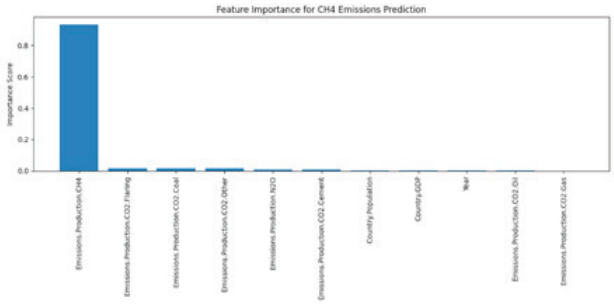

Figure 7: *Feature Importance Rankings for CH_4 Emissions Prediction using Random Forest Regression*

Figure 7 show the feature importance analysis for CH_4 emissions prediction indicates that Emissions.Production.CH4 overwhelmingly dominates the model's predictive capability, contributing over 90% of the total importance score. This result reflects the direct and expected relationship between the target variable and its own recorded production values, suggesting that historical CH_4 emission data is the primary driver in forecasting future values.

Other features, such as Emissions.Production.CO2.Flaring, Emissions.Production.CO2.Coal, and Emissions.Production. CO2.Other, show only minimal influence on CH_4 predictions, contributing marginally to the overall model performance. These minor contributions likely capture indirect correlations where certain CO_2-producing activities may also emit small quantities of CH_4, for example, through industrial processes or incomplete combustion of fossil fuels.

Socio-economic factors, including GDP and population, as well as other CO_2 sector emissions (oil, gas, and cement), have negligible importance in this model. This suggests that CH_4 emissions are largely independent of broader economic indicators and are instead strongly tied to direct measurement values. Consequently, the model's predictions for CH_4 are highly precise but heavily reliant on the availability and accuracy of historical CH_4 emission data.

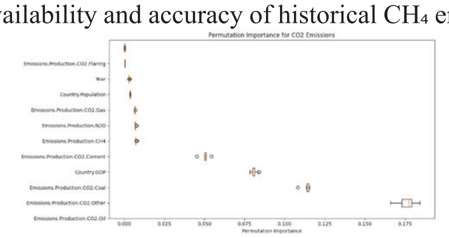

Figure 8: *Permutation Importance of Features for Predicting CO_2 Emissions.*

Figure 8 illustrates the relative importance of different features in predicting CH_4 emissions, within the framework of applying machine learning to enhance the prediction and optimization of CO_2 and CH_4 emission reduction potential. The analysis reveals that Emissions.Production.CH4 is by far the most influential predictor, accounting for the vast majority of the model's explanatory power. This dominance underscores the direct dependency of the prediction on historical CH_4 production data.

Other features, such as CO_2 emissions from flaring, coal production, and other CO_2 sources, contribute only marginally to the predictive model, indicating limited but possible indirect correlations between these activities and CH_4 emissions. These smaller influences may reflect overlapping processes in energy production and industrial activities where both CO_2 and CH_4 are released simultaneously.

Socio-economic indicators, including GDP and population, along with other CO_2-related sectors such as oil, gas, and cement, exhibit negligible predictive importance. This suggests that, unlike CO_2 emissions, CH_4 levels are less driven by broad economic factors and more directly tied to sector-specific processes. Such insights allow for more targeted methane reduction strategies, focusing on direct CH_4 emission sources rather than broad economic or indirect contributors.

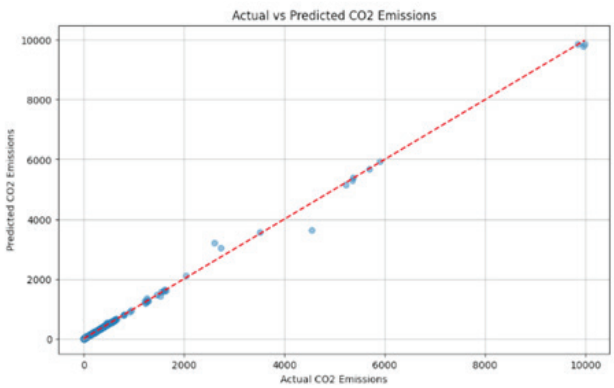

Figure 9: *Model Performance: Actual vs. Predicted CO_2 Emissions*

As shown in figure 9 the scatter plot compares model predicted CO_2 emissions with observed values; points lie tightly around the red 1:1 line, indicating an excellent fit across the full range of emissions. This visual agreement is consistent with the quantitative metrics ($R^2 \approx 0.997$ and RMSE ≈ 53.69), showing that the selected features and the Random Forest model capture the dominant drivers of CO_2 emissions with very small residual error. No clear curvature or funneling is apparent, so systematic bias (over- or under-prediction at low/high values) is minimal. A few distant points likely reflect country-year anomalies, measurement noise, or patterns underrepresented in training; these merit follow-up (e.g., data quality checks, adding sectoral context features, or robust modeling) to further tighten accuracy.

Figure 9 compares the model's predicted CO_2 emissions against the actual observed values, providing an assessment of prediction accuracy within the study on applying machine learning for enhanced CO_2 and CH_4 emission reduction potential. The close alignment of data points along the red dashed 1:1 reference line indicates that the model achieves a high degree of predictive accuracy, with minimal deviation between predicted and actual values. This strong correlation suggests that the selected features and the trained algorithm effectively capture the underlying patterns influencing CO_2 emissions, making the model reliable for forecasting and optimizing emission reduction strategies. The few visible outliers may reflect instances of unusual activity or underrepresented patterns in the dataset, warranting further investigation for model refinement.

```
CO2 Emissions Model:
Best Parameters: {'max_depth': 30, 'min_samples_split': 2, 'n_estimators': 200}
RMSE: 53.689841898414116
R² Score: 0.9971290240043951

CH4 Emissions Model:
Best Parameters: {'max_depth': None, 'min_samples_split': 2, 'n_estimators': 300}
RMSE: 2.578185330137249
R² Score: 0.9997145702405652
```

Figure 10: *Model Performance Metrics and Optimal Hyperparameters for CO_2 and CH_4 Emissions Prediction.*

The Random Forest Regression model achieved high predictive accuracy for both CO_2 and CH_4 emissions. For CO_2, the optimal parameters were a maximum depth of 30, minimum samples split of 2, and 200 estimators, resulting in an RMSE of 53.69 and an R^2 score of 0.9971, indicating an excellent fit between predicted and actual values. For CH_4, the best model used no limit on tree depth, a minimum samples split of 2, and 300 estimators, yielding an RMSE of 2.58 and an R^2 score of 0.9997, reflecting near perfect predictive performance. These results demonstrate the model's robustness and ability to capture complex patterns in emission data as shown in Figure 10.

Table: Machine Learning Results for CO_2 and CH_4 Emission Prediction

Figure	Analysis	Key Findings	Implications
Figure 4	Time Series Trends of CO_2 Emissions (Top 5 countries, 1990–2018)	1. China: sharp increase post-2000 due to industrialization & coal use. 2. USA: high but stable, slight decline post-2007 (clean energy adoption). 3. India: steady rise (industrial/energy demand). 4. Russia: drop in 1990s, later stabilized. 5. Japan: flat with minor fluctuations.	Country specific emission drivers highlight the need for tailored reduction strategies.
Figure 5	Correlation Heatmap (Socio-economic & sectoral factors with CO_2 & CH_4)	1. CO_2 strongly correlated with GDP (0.94), coal (0.95), oil (0.94). 2. CH_4 correlated with population (0.84) & N_2O (0.92).	Strong interdependencies indicate that multivariate ML models can capture cross gas and socio-economic effects.
Figure 6	Feature Importance CO_2 (Random Forest)	1. Top drivers: CO_2 (Other, Oil, Coal) → >75% of importance. 2. Moderate: GDP, N_2O. 3. Minimal: Population, gas, flaring.	Sector specific CO_2 sources dominate total emissions prediction.

Figure 7	Feature Importance CH_4 (Random Forest)	1. CH_4 production alone accounts for > 90%. 2. Minor roles: CO_2 flaring, coal, other sources. 3. Negligible: GDP, population, oil/gas sectors.	CH_4 emissions are highly dependent on direct measurement data, less on socio-economics.
Figure 8	Permutation Importance CH_4	1. CH_4 production overwhelmingly dominant. 2. Minor: CO_2 flaring, coal, other sources. 3. Negligible: GDP, population, oil, gas, cement.	Confirms direct dependency on historical CH_4 values; useful for targeted methane reduction.
Figure 9	Model Performance (CO_2: Actual vs Predicted)	1. $R^2 \approx 0.997$, RMSE $\approx$ 53.69. 2. Predictions tightly align with observed values. 3. Few anomalies (data noise, underrepresented cases).	Random Forest captures main CO_2 drivers with very high accuracy.
Figure 10	Model Performance Metrics & Hyperparameters	1. CO_2 model: Depth=30, Estimators = 200 → R^2=0.9971, RMSE=53.69. 2. CH_4 model: Unlimited depth, Estimators=300 → R^2=0.9997, RMSE=2.58.	Both models show exceptional robustness & predictive

Conclusion

This study successfully applied a data-driven machine learning framework to enhance the prediction and optimization of CO_2 and CH_4 emission reduction potential across multiple countries. Using Random Forest Regression, the models achieved exceptionally high predictive accuracy ($R^2 \approx 0.997$, RMSE $\approx$ 53.69), closely aligning predicted emissions with observed values and demonstrating minimal systematic bias. Feature importance analysis revealed that oil, coal, and other CO_2 sources are the most influential drivers of total emissions, while GDP and cement production also play notable roles. Correlation analysis further highlighted the strong interdependence between greenhouse gas outputs and socio-economic as well as sectoral factors, underscoring the need for integrated policy measures.

The novelty of this work lies in combining high resolution multi sectoral and socio-economic data with advanced machine learning techniques to simultaneously forecast emissions and identify priority intervention points. This dual capability allows not only for accurate prediction but also for targeted scenario simulation enabling policymakers to test the impact of potential measures, such as fuel switching or economic adjustments, before implementation. By bridging predictive accuracy with actionable insights, this approach provides a robust, scalable, and evidence-based pathway for accelerating global decarbonization and methane mitigation strategies.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions:

Ajayi Abiola Samuel – Conceptualization; Methodology; Data curation; Machine learning model development; Formal analysis; Software implementation; Visualization; Writing – original draft; Project administration; Correspondence; Validation.

Shokenu Emmanuel Segun – Methodology; Statistical analysis; Data interpretation; Technical review of model framework; Writing – review & editing.

Ajayi Eniola Isaac – Data preprocessing; Exploratory data analysis; Feature engineering; Writing – review & editing; Assistance with figures and tables.

Godwin Iheuwa – Literature review; Theoretical framework support; Critical revision of related works; Writing – review & editing.

Ayoola David Bodude – Domain expertise on emission sources and energy processes; Interpretation of results; Validation of sector-based emission insights; Writing – editing.

Brian Cobb Hanrahan – Oversight and supervision; Industry insights on energy, aerospace, and emission reduction technologies; Interpretation and policy relevance analysis; Final manuscript review.

All authors reviewed and approved the final version of the manuscript and agree to its submission.

AI statement: The authors stated that generative AI tools (specifically ChatGPT) were used only for language polishing and formatting assistance. The authors reviewed and verified all content to ensure accuracy and integrity of the scientific work.

Reference

Filonchyk, M., Peterson, M. P., Zhang, L., Hurynovich, V. and He, Y. (2024) 'Greenhouse gas emissions and global climate change: Examining the influence of CO_2, CH_4 and N_2O', *Science of the Total Environment*, vol. 935, p. 173359. doi:10.1016/j.scitotenv.2024.173359.

Pulles, T. and Van Amstel, A. (2010) 'An overview of non-CO_2 greenhouse gases', *Journal of Integrative Environmental Sciences*, vol. 7, no. S1, pp. 3–19. doi:10.1080/1943815X.2010.505241.

Brander, M. and Davis, G. (2012) 'Greenhouse gases, CO_2, CO_2e and carbon: What do all these terms mean?', *Econometrica White Papers*, pp. 2–3.

Alagade, A. and Sahu, M. (2025) 'Satellite-based assessment and forecasting of greenhouse gas (GHG) concentrations in Indian megacities using advanced statistical methods', *Environmental Science and Pollution Research*, pp. 1–19. doi:10.1007/s11356-025-36583-1.

Dang, P., Zhang, M., Chen, X., Loreau, M., Duffy, J. E., Li, X. E. et al. (2024) 'Plant diversity decreases greenhouse gas emissions by increasing soil and plant carbon storage in terrestrial ecosystems', *Ecology Letters*, vol. 27, no. 7, e14469. doi: 10.1111/ele.14469.

Zhang, Q., Li, J., Li, L. and Yu, J. (2025) 'Zeolite-based materials for greenhouse gas capture and conversion', *Science*

China Chemistry, vol. 68, no. 5, pp. 1703–1716. doi: 10.1007/s11426-024-2287-6.

Faraday, E. T. and Oluwabunmi, O. A. (2024) 'Greenhouse gas levels (CH_4 and CO_2) in Lagos State and Oyo State, Nigeria', *Discover Environment*, vol. 2, no. 1, p. 4. doi:10.1007/s44274-023-00026-w.

Zhang, X., Huang, T., Shan, T., Yuan, Q., Li, J., Wu, Q. and Zhang, P. (2024) 'A comprehensive review of molecular dynamics simulation on the replacement characteristics and mechanism of CO_2–CH_4 hydrate in porous media systems', *Greenhouse Gases: Science and Technology*, vol. 14, no. 4, pp. 695–710. doi:10.1002/ghg.2292.

Wu, G., Feng, G., Li, Y., Ling, T., Peng, X., Su, Z. and Zhao, X. (2024) 'A review of thermal energy management of diesel exhaust after-treatment systems technology and efficiency enhancement approaches', *Energies*, vol. 17, no. 3, p. 584. doi:10.3390/en17030584.

Bakhchin, D., Ravi, R., Douadi, O., Faqir, M. and Essadiqi, E. (2024) 'Integrated catalytic systems for simultaneous NO_x and PM reduction: A comprehensive evaluation of synergistic performance and combustion waste energy utilization', *Environmental Science and Pollution Research*, vol. 31, no. 34, pp. 46840–46857. doi:10.1007/s11356-024-34287-6.

Li, T., Yue, X. G., Qin, M. and Norena-Chavez, D. (2024) 'Towards Paris Climate Agreement goals: The essential role of green finance and green technology', *Energy Economics*, vol. 129, p. 107273. doi:10.1016/j.eneco.2023.107273.

Kreibich, N. (2024) 'Toward global net zero: The voluntary carbon market on its quest to find its place in the post-Paris climate regime', *Wiley Interdisciplinary Reviews: Climate Change*, vol. 15, no. 5, e892. doi:10.1002/wcc.892.

Park, S., Moon, K.-J., Eom, H.-J., Yi, S.-M., Kim, Y., Kim, M., Rim, D. and Lee, Y. S. (2025) 'Machine learning-based prediction of ambient CO_2 and CH_4 concentrations with high temporal resolution in Seoul metropolitan area', *Environmental Pollution*, vol. 345, p. 126362. doi:10.1016/j.envpol.2025.126362.

Essamlali, I., Nhaila, H. and El Khaili, M. (2024) 'Supervised machine learning approaches for predicting key pollutants and for the sustainable enhancement of urban air quality: A systematic review', *Sustainability*, vol. 16, no. 3, p. 976. doi:10.3390/su16030976.

Hua, J., Wang, R., Hu, Y., Chen, Z., Chen, L., Osman, A. I., Farghali, M., Huang, L., Feng, J., Wang, J., Zhang, X., Zhou, X. and Yap, P.-S. (2025) 'Artificial intelligence for calculating and predicting building carbon emissions: A review', *Environmental Chemistry Letters*, vol. 23, no. 3, pp. 783–816. doi:10.1007/s10311-024-01799-z.

Du, X., Khan, M. N. and Thakur, G. C. (2025) 'Machine learning in carbon capture, utilization, storage, and transportation: A review of applications in greenhouse gas emissions reduction', *Processes*, vol. 13, no. 4, p. 1160. doi:10.3390/pr13041160.

Gan, N. and Zhao, S. (2024) 'Global greenhouse gas reduction forecasting via machine learning model in the scenario of energy transition', *Journal of Environmental Management*, vol. 371, p. 123309. doi:10.1016/j.jenvman.2024.123309.

Shafaghat, A. (2025) 'Enhancing carbon capture and CO_2 reduction processes using machine learning and AI technologies to improve biomedical outcomes', *Metallurgical and Materials Engineering*, vol. 31, no. 2, pp. 129–146. doi:10.63278/mme.v31i2.1851.

Kothale, A. and Sadgir, P. (2025) 'Application of artificial intelligence and machine learning with international guidelines for greenhouse gas reduction in wastewater treatment', *International Journal of Environmental Science and Technology*, pp. 1–19. doi:10.1007/s13762-025-06651-7.

Ojadi, J. O., Onukwulu, E., Odionu, C. and Owulade, O. (2023) 'AI-driven predictive analytics for carbon emission reduction in industrial manufacturing: A machine learning approach to sustainable production', *International Journal of Multidisciplinary Research and Growth Evaluation,* vol. 4, no. 1, pp. 948–960. doi:10.54660/IJMRGE.2023.4.1.948-960.

Xue, H., Wang, G., Gui, X., Gong, H., Li, X. and Du, F. (2024) 'A novel multidimensional hybrid machine learning model for CO_2 injection to separate coalbed methane: Comprehensive prediction of methane diffusion rate, production volume, and CO_2 sequestration', *Energy & Fuels*, vol. 38, no. 18, pp. 17525–17540. doi:10.1021/acs.energyfuels.4c02621.

Xue, H., Wang, G., Li, X. and Du, F. (2024) 'Predictive combination model for CH_4 separation and CO_2 sequestration with CO_2 injection into coal seams: VMD-STA-BiLSTM-ELM hybrid neural network modeling', *Energy*, vol. 313, p. 133744. doi:10.1016/j.energy.2024.133744.

Ehinmowo, A. B., Nwaneri, B. I. and Olaide, J. O. (2025) 'Predictive modeling of hydrogen production and methane conversion from biomass-derived methane using machine learning and optimisation techniques', *Next Energy*, vol. 7, p. 100229. doi:10.1016/j.nxener.2024.100229.

Wang, H., Li, Y., Xuan, X., Wang, K., Yao, Y. F. and Pan, L. (2025) 'Machine learning accelerated discovery of covalent organic frameworks for environmental and energy applications', *Environmental Science and Technology*, vol. 59, no. 13, pp. 6361–6378. doi:10.1021/acs.est.5c00390.

Zhang, D. and Feng, E. (2024) 'Quantitative assessment of regional carbon neutrality policy synergies based on deep learning', *Journal of Advanced Computing Systems*, vol. 4, no. 10, pp. 38–54. doi:10.69987/JACS.2024.41004.

Ajayi, A. S., Kim, S. and 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*, vol. 32, p. 15. doi:10.1007/s44189-024-00060-0.

Kim, Y., Samuel, A. A. and 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*, vol. 35, no. 11, pp. 557–565. doi:10.6110/KJACR.2023.35.11.557.

Ehiziojie, L., Adebara, O. A., Okoro, U. D., Ajayi, S. and Adeloye, K. (2024) 'Mitigating the impact of scope change on project cost performance in the global oil and gas industry: Innovative strategies and sustainable solutions from the Nigerian experience', in *World Conference on Future Innovations and Sustainable Solutions*. Futurity Research Publishing. doi:10.5281/zenodo.13993544.